

## PROMOTER VARIANTS OF THE ALPHA-7 NICOTINIC ACETYLCHOLINE RECEPTOR

This invention was made with government support under National Institutes of Health  
5 Grants DA09457, DA12231, AG00029, MH36321, and MH44212, and the Veterans Affairs  
Medical Research Service. The Government has certain rights in the invention.

This is a Continuation-In-Part of co-pending Appln. No. 08/956,518, filed on October 23,  
1997

### FIELD OF THE INVENTION

The present invention is related to the alpha-7 neuronal nicotinic acetylcholine receptor  
gene. In particular, the present invention provides the human alpha-7 gene. In addition, the  
present invention provides methods and compositions for the diagnosis and treatment of  
15 schizophrenia.

### BACKGROUND OF THE INVENTION

Schizophrenia is the most common chronic psychotic disorder of humans, affecting  
approximately one percent of the population worldwide (Eaton, Epidemiol Rev, 7:105, 1985).  
20 The mean lifetime risk of schizophrenia has been estimated at one percent (Eaton, *supra*, 985).  
As the onset of disease usually occurs early in life, and results in serious chronic impairment of  
cognition, behavior, and emotional state, schizophrenia is a major social problem in terms of  
cost, lost potential and productivity, and family stress. Furthermore, estimates indicate that the  
mortality of schizophrenic patients is twice that of the general population (Tsuang *et al.*, Arch  
25 Gen Psychiat, 36:1295, 1979). The medical care of schizophrenic patients also presents a  
significant challenge, as the patients are often unable to provide an accurate medical history, and  
have difficulty complying with medical treatment.

The essential features of schizophrenia are the presence of psychotic symptoms during  
some phase of the illness, a chronic course, and deterioration in function. However, no  
30 combination of signs or symptoms is truly pathognomic of the disease. The DSM-IV criteria for  
schizophrenia (*See*, Hyman, "Schizophrenia," in Dale and Federman (eds.), *Scientific American*

*Medicine*, New York, New York, 13 VII:1-5, 1994), requires a duration of at least six months, and a deterioration in function. Psychotic symptoms typically exhibited by schizophrenia patients include disturbances in perception, abnormalities in thought content, and abnormalities in the form of thought. Perceptual disturbances typically consist of hallucinations and illusions.

5 The course of schizophrenia is variable, although it is generally characterized by periods with exacerbation of psychotic symptoms, followed by periods of remission. Over a period of years, social and cognitive function usually deteriorates. Suicide attempts and depression are common. As measured by frequency and severity of relapses, continuing symptoms, and overall functioning, approximately 80% of schizophrenics have a poor outcome (Breier *et al.*, Arch Gen  
10 Psychiat, 48:239, 1991).

Although family, twin, and adoption studies indicate that schizophrenia has a significant genetic component, these studies also show that the inheritance of schizophrenia is complex, involving an uncertain mode of transmission, incomplete penetrance, and probable genetic heterogeneity (Risch, Genet Epidemiol, 7:3, 1990; and Tsuang, Brit J Psychiat, 163:299, 1993).

15 Linkage studies using schizophrenia and related psychiatric cases as phenotypes have found possible loci for schizophrenia at various chromosomal sites in subsets of families (Pulver *et al.*, Am J Med Genet, 54:44, 1994; Coon *et al.*, Am J Med Genet, 54:12, 1994; Wang *et al.*, Nature Genet, 10:41, 1995; and Silverman *et al.*, Am J Med Genet, 67:162, 1996). However, the findings do not account completely for the inheritance of schizophrenia, nor do they delineate  
20 which aspects of this multifactorial illness might be influenced by a specific locus.

A variety of psychiatric disorders may mimic schizophrenia and the symptoms of many disorders are similar. Thus, diagnosis has been based on the course of illness (for example, acute onset and episodic course in mania, compared with an insidious onset and chronic course in schizophrenia). In addition to schizophrenia, psychotic symptoms may also occur as a result of  
25 metabolic disturbances, structural brain lesions, other medical conditions, or drug toxicity. Thus, the differential diagnosis of schizophrenia must take into consideration such medical conditions as central nervous system neoplasm, hyperviscosity syndromes (*i.e.*, due to hematologic malignancy), paraneoplastic syndromes, anoxia and postanoxic encephalopathy, hypertensive encephalopathy, AIDS encephalopathy, encephalitis, meningitis, brain abscess, Lyme disease,  
30 neurosyphilis, acute intermittent porphyria, Addison's disease, Cushing's disease, hepatic encephalopathy, hypocalcemia, hypercalcemia, hypoglycemia, hypothyroidism,

hyperthyroidism, Alzheimer's disease, complex partial seizures, Huntington's disease, multiple sclerosis, stroke, Wilson's disease, folic acid deficiency, pellagra, vitamin B<sub>12</sub> deficiency, and lupus cerebritis. Some drugs, such as alcohol, high-dose cocaine, high-dose amphetamines, marijuana, phencyclidine (PCP), hallucinogens, sedative-hypnotics, meperidine, non-steroidal  
5 anti-inflammatory drugs, pentazocine and other opiate mixed agonist-antagonists, anti-tuberculosis drugs (*e.g.*, cycloserine, isoniazid, rifampin), other antimicrobials, anticholinergic anti-parkinsonians, anti-histamines (*e.g.*, diphenhydramine), atropine and derivatives, cyclic antidepressants, low-potency antipsychotic drugs (*e.g.*, thioridazine and clozapine), meclizine, scopolamine, anti-arrhythmic (*e.g.*, amiodarone, digitalis, and procainamide), captopril,  
10 amantadine, D<sub>2</sub> dopamine receptor antagonists (*e.g.*, bromocriptine, and pergolide), levodopa, estrogens, testosterone, glucocorticoids and adrenocorticotrophic hormone (ACTH), thyroid replacement overdose, cimetidine, ranitidine, dextroamphetamine, methylphenidate, and over-the-counter decongestants (*e.g.*, pseudoephedrine), diet pills, and pep pills, are commonly associated with psychotic symptoms.

15 Treatment of schizophrenic patients usually involves the use of anti-psychotic drugs (*e.g.*, haloperidol, haloperidol-like drugs, and atypical neuroleptics such as clozapine), maintenance of a safe, predictable environment, and supportive psychotherapy to improve social and coping skills of patients. Stress reduction also appears to prevent relapses. While these drugs are useful in treating the symptoms of schizophrenia, there are also problems associated with their use. For  
20 example, the use of clozapine is complicated by the idiosyncratic occurrence of agranulocytosis, necessitating weekly monitoring of the white blood cell counts of patients taking this drug (*See*, Hyman, *supra*, 1994).

Despite advances in treatment and diagnostic methods, there remains a need for methods to diagnose and treat schizophrenic patients. Indeed, methods to diagnose and screen large  
25 populations for genetic component(s) associated with schizophrenia, as well as other psychoses are needed in order to provide reliable diagnoses that are not dependent upon the multifactorial criteria presently in use. Improved methods of treatment are also needed, including drugs and other therapeutics that do not have the side effects and other undesirable properties associated with the currently used drugs.

## SUMMARY OF THE INVENTION

The present invention is related to the alpha-7 neuronal nicotinic acetylcholine receptor gene. In particular, the present invention provides the human alpha-7 gene. In addition, the present invention provides methods and compositions for the diagnosis and treatment of schizophrenia.

In one embodiment, the present invention provides an isolated nucleotide sequence encoding at least a portion of the human alpha-7 nicotinic receptor, wherein the sequence is selected from the group consisting of SEQ ID NOS:84-103. In an alternative embodiment, the present invention provides an isolated peptide sequence encoded by the isolated nucleotide sequence, wherein the nucleotide sequence is selected from the group consisting of SEQ ID NOS:84-103. In another embodiment, the nucleotide sequence further comprises 5' and 3' flanking regions. In yet another embodiment, the nucleotide sequence further comprises intervening regions. In a further embodiment, the present invention provides an isolated polynucleotide sequence comprising a combination of two or more nucleotide sequences, wherein the nucleotide sequences are selected from the group consisting of SEQ ID NOS:84-103. It is not intended that the combination comprise any particular number or order of these nucleotide sequences, nor is it intended that the combination be limited to the inclusion of any particular nucleotide sequence.

In another embodiment, the present invention provides vectors comprising a nucleotide sequence, wherein the nucleotide sequence comprises at least one nucleotide sequence selected from the group consisting of SEQ ID NOS:84-103. In another embodiment, the present invention provides a host cell transformed with a vector comprising a nucleotide sequence, wherein the nucleotide sequence comprises at least one nucleotide sequence selected from the group consisting of SEQ ID NOS:84-103. In one embodiment, the host cell is selected from the group consisting of bacteria, yeast, amphibian, and mammalian cells. In one preferred embodiment, the host cell is a human cell. In an alternative preferred embodiment, the host cell is a cell line, while in another preferred embodiment, the host cell is contained within an animal.

The present invention also provides a first polynucleotide sequence comprising at least fifteen nucleotides, which hybridizes under stringent conditions to at least a portion of a second polynucleotide sequence, wherein the second polynucleotide sequence is selected from the polynucleotide sequences selected from the group consisting of SEQ ID NOS:84-103..

The present invention also provides methods for detection of a polynucleotide encoding alpha-7 protein in a biological sample suspected of containing the polynucleotide encoding alpha-7, comprising the step of hybridizing at least a portion of a polynucleotide sequence selected from the group consisting of SEQ ID NOS:9-11, and 84-103, to nucleic acid of the biological sample to produce an hybridization complex. In one embodiment, the method further comprises the step of detecting the hybridization complex, wherein the presence of the complex correlates with the presence of a polynucleotide encoding alpha-7 in the biological sample. In another embodiment, the biological sample is a sample selected from the group consisting of brain tissue and blood. In one preferred embodiment, the biological sample is from a human. In yet another embodiment, the human is suspected of suffering from a condition selected from the group consisting of schizophrenia, small cell lung carcinoma, breast cancer, and nicotine-dependent illness. In yet another preferred embodiment of the method, the nucleic acid of the biological sample is amplified by the polymerase chain reaction prior to hybridization.

The present invention also provides methods for amplification of nucleic acid from a sample suspected of containing nucleic acid encoding alpha-7, comprising the steps of: providing a test sample suspected of containing amplifiable nucleic acid encoding alpha-7; isolating the amplifiable nucleic acid from the test sample; combining the amplifiable nucleic acid with amplification reagents, and at least two primers selected from the group consisting of primers having the nucleic acid sequence set forth in SEQ ID NOS:1-8, and 12-83 to form a reaction mixture; and combining the reaction mixture with an amplification enzyme under conditions wherein the amplifiable nucleic acid is amplified to form amplification product. In one embodiment, the method further comprises the step of detecting the amplification product. In an alternative embodiment, the detecting is accomplished by hybridization of the amplification product with a probe having the nucleic acid sequence is selected from group of the sequences set forth in SEQ ID NO:9-11. In one preferred embodiment, the test sample is a sample selected from the group consisting of brain tissue and blood. In an alternative preferred embodiment, the test sample is from a human. In yet another embodiment, the human is suspected of suffering from a condition selected from the group consisting of schizophrenia, small cell lung carcinoma, breast cancer, and nicotine-dependent illness.

The present invention also provides methods for amplification of nucleic acid from a sample suspected of containing nucleic acid encoding alpha-7 comprising the steps of: providing

a test sample suspected of containing amplifiable nucleic acid encoding alpha-7; isolating the amplifiable nucleic acid from the test sample; combining the amplifiable nucleic acid with amplification reagents, and a first primer set comprising at least two primers selected from the group consisting of the sequences set forth in SEQ ID NOS:65-70, to form a first reaction mixture; combining the reaction mixture with an amplification enzyme under conditions wherein the amplifiable nucleic acid is amplified to form a first amplification product; combining the first reaction mixture with amplification reagents, and a second primer set comprising at least two primers selected from the group consisting of the sequences set forth in SEQ ID NOS:57-59, 61, 63, 67, and 73-75, to form a second reaction mixture; combining the second reaction mixture with an amplification enzyme under conditions wherein the amplifiable nucleic acid is amplified to form a second amplification product; and detecting the first or second amplification product.

In one preferred embodiment of the method, the detecting comprises hybridizing the amplification product with a probe having a nucleic acid sequence selected from the group consisting of the nucleic acid sequence set forth in SEQ ID NOS:9-11. In yet another embodiment, the test sample is a sample selected from the group consisting of brain tissue and blood. In another preferred embodiment of the method, the test sample is from a human. In a further embodiment, the human is suspected of suffering from a condition selected from the group consisting of schizophrenia, small cell lung carcinoma, breast cancer, and nicotine-dependent illness.

Additionally, the present invention provides methods of identifying individuals predisposed schizophrenia comprising: providing a nucleic acid from a human subject; wherein the nucleic acid comprises an  $\alpha 7$  allele; detecting the presence of at least one polymorphism within the  $\alpha 7$  allele; and correlating the presence of the at least one polymorphism with a predisposition to schizophrenia. In some embodiments the at least one polymorphism comprises one or more of a -241 A to G substitution, a -194 G to C substitution, a -191 G to A substitution, a -190 G insertion, a -180 G to C substitution, a -178 CGGGGG insertion, a -178 G deletion, a -166 C to T substitution, a -143 G to A substitution, a -140 CGGG insertion, a -93 C to G substitution, a -92 G to A substitution, a -86 C to T substitution, and a -46 G to T substitution. In other embodiments the at least one polymorphism comprises two or more polymorphisms. In some preferred embodiments, the at least one polymorphism comprises a promoter polymorphism that contributes to reduced  $\alpha 7$  transcription. The present invention

provides methods wherein the detecting step is accomplished using at least one technique selected from the group consisting of polymerase chain reaction, heteroduplex analysis, single stand conformational polymorphism analysis, denaturing high performance liquid chromatography, ligase chain reaction, comparative genome hybridisation, Southern blotting and sequencing. In some embodiments, the nucleic acid from the subject is derived from a sample selected from the group consisting of a biopsy material and blood. Moreover embodiments are provided which further comprise step d) providing a diagnosis to the subject based on the presence or absence of the at least one polymorphism. In preferred embodiments, the diagnosis differentiates schizophrenia from other forms of mental illness.

The present invention also provides kits for determining if a subject is predisposed to schizophrenia, comprising: at least one reagent suitable for use in specifically detecting at least one polymorphism in an  $\alpha 7$  allele; and instructions for determining whether a subject is predisposed to schizophrenia. In some embodiments, the at least one polymorphism comprises one or more of a -241 A to G substitution, a -194 G to C substitution, a -191 G to A substitution, a -190 G insertion, a -180 G to C substitution, a -178 CGGGGG insertion, a -178 G deletion, a -166 C to T substitution, a -143 G to A substitution, a -140 CGGG insertion, a -93 C to G substitution, a -92 G to A substitution, a -86 C to T substitution, and a -46 G to T substitution. In preferred embodiments, the at least one polymorphism comprises a promoter polymorphism that contributes to reduced  $\alpha 7$  transcription. The present invention further provides embodiments in which the at least one reagent comprises a nucleic acid probe that hybridizes under stringent conditions to a nucleic acid sequence selected from the group consisting of the coding strand of the  $\alpha 7$  gene, and the noncoding strand of the  $\alpha 7$  gene. In some preferred embodiments, the at least one reagent comprises a sense primer and an antisense primer flanking the at least one polymorphism in the  $\alpha 7$  allele. In a subset of these, at least one of the primers comprises a fluorescent tag. Moreover, in some embodiments, the instructions comprise instructions required by the United States Food and Drug Administration for use in *in vitro* diagnostic products.

Also provided by the present invention are methods of screening compounds, comprising: providing: i) at least one cell comprising an  $\alpha 7$  allele with at least one polymorphism, and ii) one or more test compounds; and contacting the at least one cell with the test compound; and detecting a change in  $\alpha 7$  expression in the at least one cell in the presence of the test compound

relative to the absence of the test compound. In some embodiments the detecting comprises detecting  $\alpha 7$  mRNA, while in others the detecting comprises detecting  $\alpha 7$  polypeptide. In preferred embodiments, the cell is a neuroblastoma cell. In other preferred embodiments, the test compound comprises a drug. Moreover, in particularly preferred embodiments, the at least one polymorphism comprises a promoter polymorphism that contributes to reduced  $\alpha 7$  transcription.

In alternative embodiments, the present invention provides methods of screening compounds, comprising: providing: i) at least one cell comprising an  $\alpha 7$  promoter in operable combination with a reporter gene, wherein said  $\alpha 7$  promoter comprises at least one polymorphism, and ii) one or more test compounds; and contacting the at least one cell with the test compound; and detecting a change in expression of the reporter gene in the at least one cell in the presence of the test compound relative to the absence of the test compound. In some embodiments the detecting comprises detecting reporter gene mRNA or polypeptide, while in others the detecting comprises detecting reporter gene function. In preferred embodiments, the cell is a neuroblastoma cell. In other preferred embodiments, the test compound comprises a drug. In an exemplary embodiment, the reporter gene is the firefly luciferase gene.

Furthermore, the present invention provides methods of identifying individuals predisposed to schizophrenia, comprising: providing a nucleic acid sample from a subject, the sample containing an  $\alpha 7$  allele; correlating the identity of the  $\alpha 7$  allele with a predisposition to schizophrenia. In some embodiments, the identity of the  $\alpha 7$  allele is determined using at least one technique selected from the group consisting of polymerase chain reaction, heteroduplex analysis, single stand conformational polymorphism analysis, denaturing high performance liquid chromatography, ligase chain reaction, comparative genome hybridisation, Southern blotting and sequencing. In preferred embodiments, the nucleic acid sample from the subject is selected from the group consisting of a biopsy material and blood. Moreover embodiments are provided which further comprise step c) providing a diagnosis to the subject based on the identity of the  $\alpha 7$  allele.

The present invention also provides methods for producing anti- $\alpha 7$  antibodies (including, but not limited to antibodies directed against peptides comprising  $\alpha 7$ ), comprising, exposing an animal having immunocompetent cells to an immunogen comprising at least an antigenic portion of  $\alpha 7$  protein, under conditions such that immunocompetent cells produce antibodies directed

against the portion of  $\alpha 7$ . In preferred embodiments, the  $\alpha 7$  peptide or protein is human  $\alpha 7$ . In one embodiment, the method further comprises the step of harvesting the antibodies. In an alternative embodiment, the method comprises the step of fusing the immunocompetent cells with an immortal cell line under conditions such that a hybridoma is produced. In other  
5   embodiments, the immunogen comprises a fusion protein.

The present invention also provides methods for detecting abnormal  $\alpha 7$  expression comprising the steps of: a) providing a sample suspected of containing test  $\alpha 7$ ; and a control containing a quantitated  $\alpha 7$ ; and b) comparing the test  $\alpha 7$  in the sample with the quantitated  $\alpha 7$  in the control to determine the relative concentration of the test  $\alpha 7$  in the sample. In one  
10   embodiment of the method, the control contains a higher concentration of quantitated  $\alpha 7$  than the concentration of the test  $\alpha 7$  in the sample. Thus, the methods are capable of identifying samples (e.g., patient samples) with reduced  $\alpha 7$  protein. The methods also provide means to detect samples that contain a normal amount of the  $\alpha 7$  protein. In addition, the methods may be conducted using any suitable means to determine the relative concentration of  $\alpha 7$  in the test and  
15   control samples, including but not limited to the means selected from the group consisting of Western blot analysis, Northern blot analysis, Southern blot analysis, denaturing polyacrylamide gel electrophoresis, reverse transcriptase-coupled polymerase chain reaction, enzyme-linked immunosorbent assay, radioimmunoassay, and fluorescent immunoassay. Thus, the methods may be conducted to determine the presence of  $\alpha 7$  in the genome of the animal source of the test  
20   sample, or the expression of  $\alpha 7$  (mRNA or protein), as well as detect the presence of abnormal or mutated  $\alpha 7$  proteins or gene sequences in the test samples.

In one preferred embodiment, the presence of  $\alpha 7$  is detected by immunochemical analysis. For example, the immunochemical analysis can comprise detecting binding of an antibody specific for an epitope of  $\alpha 7$ . In another preferred embodiment of the method, the  
25   antibody comprises polyclonal antibodies, while in another preferred embodiment, the antibody is comprises monoclonal antibodies.

The antibodies used in the methods invention may be prepared using various immunogens. In one embodiment, the immunogen is a human  $\alpha 7$  peptide to generate antibodies that recognize human  $\alpha 7$ . Such antibodies include, but are not limited to polyclonal,  
30   monoclonal, chimeric, single chain, Fab fragments, and a Fab expression library.

Various procedures known in the art may be used for the production of polyclonal antibodies to  $\alpha 7$  (e.g., human  $\alpha 7$ ). For the production of antibody, various host animals can be immunized by injection with the peptide corresponding to the human  $\alpha 7$  epitope including but not limited to rabbits, mice, rats, sheep, goats, etc. In a preferred embodiment, the peptide is conjugated to an immunogenic carrier (e.g., diphtheria toxoid; bovine serum albumin, BSA; or keyhole limpet hemocyanin, KLH). Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (Bacille Calmette-Guerin) and *Corynebacterium parvum*.

For preparation of monoclonal antibodies directed against  $\alpha 7$ , any technique that provides for the production of antibody molecules by continuous cell lines in culture may be used (See, e.g., Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). These include but are not limited to: the hybridoma technique originally developed by Köhler and Milstein (Köhler and Milstein, *Nature* 256:495-497, 1975); the trioma technique; the human B-cell hybridoma technique (See e.g., Kozbor *et al.*, *Immunol Today*, 4:72, 1983); and the EBV-hybridoma technique (Cole *et al.*, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96, 1985).

In an additional embodiment of the invention, monoclonal antibodies can be produced in germ-free animals utilizing recent technology (See e.g., PCT/US90/02545). According to the invention, human antibodies may be used and can be obtained by using human hybridomas (Cote *et al.*, *Proc Natl Acad Sci USA*, 80:2026-2030, 1983), or by transforming human B cells with EBV virus *in vitro* (Cole *et al.*, *supra*, 1985).

According to the invention, techniques described for the production of single chain antibodies (U.S. Patent 4,946,778; herein incorporated by reference) can be adapted to produce  $\alpha 7$  single chain antibodies. An additional embodiment of the invention utilizes the techniques described for the construction of Fab expression libraries (Huse *et al.*, *Science*, 246:1275-1281, 1989) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for  $\alpha 7$ .

Antibody fragments that contain the idiotype (antigen binding region) of the antibody molecule can be generated by known techniques. For example, such fragments include but are not limited to: the F(ab')<sub>2</sub> fragment which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments which can be generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragment, and the Fab fragments which can be generated by treating the antibody molecule with papain and a reducing agent.

In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art including but not limited to radioimmunoassay, enzyme-linked immunosorbent assay, "sandwich" immunoassay, gel diffusion precipitin reactions, immunodiffusion assays, *in situ* immunoassays (using colloidal gold, enzyme or radioisotope labels, for example), Western Blots, precipitation reactions, agglutination assays (*e.g.*, gel agglutination assays, hemagglutination assays, etc.), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc.

In one embodiment, antibody binding is detected by detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labeled. Many means are known in the art for detecting binding in an immunoassay and are within the scope of the present invention. (As is well known in the art, the immunogenic peptide should be provided free of the carrier molecule used in any immunization protocol. For example, if the peptide was conjugated to KLH, it may be conjugated to BSA, or used directly, in a screening assay.)

The foregoing antibodies can be used in methods known in the art relating to the localization and structure of  $\alpha 7$  (*e.g.*, for Western blotting), measuring levels thereof in appropriate biological samples, etc. The antibodies can be used to detect  $\alpha 7$  in a biological sample from an individual. The biological sample can be a biological fluid, such as but not limited to, blood, serum, plasma, cerebrospinal fluid (CSF), and the like, containing cells. In particular,  $\alpha 7$  can be detected from cellular sources, such as, but not limited to, brain tissue.

The biological samples can then be tested directly for the presence of  $\alpha 7$  using an appropriate strategy (*e.g.*, ELISA or radioimmunoassay) and format (*e.g.*, microwells, dipstick for instance as described in International Patent Publication WO 93/03367, etc.). Alternatively, proteins in the sample can be size separated (*e.g.*, by polyacrylamide gel electrophoresis, PAGE,

in the presence or absence of sodium dodecyl sulfate, SDS), and the presence of  $\alpha 7$  detected by immunoblotting (Western blotting). Immunoblotting techniques are generally more effective with antibodies generated against a peptide corresponding to an epitope of a protein, and hence, are particularly suited to the present invention.

5           The foregoing explanations of particular assay systems are presented herein for purposes of illustration only, in fulfillment of the duty to present an enabling disclosure of the invention. It is to be understood that the present invention contemplates a variety of immunochemical assay protocols within its spirit and scope.

10           In some preferred aspects, genomic DNA or mRNA is amplified by PCR, and the amplified DNA is tested for the presence of mutation(s). PCR amplification is well known in the art (Cameron *et al.*, Science, 257:383-387, 1992; Saksela *et al.*, Proc Natl Acad Sci USA, 91:1104-1108, 1994). For example, mRNA can be detected by reverse transcriptase-initiated PCR (*See, e.g.*, Saksela *et al.*, J Virol, 67:7423-27, 1993). PCR can be carried out (*e.g.*, by use of a Perkin-Elmer Cetus thermal cycler and *Taq* polymerase, *e.g.*, Gene Amp®, Boehringer  
15 Mannheim). The amplified PCR products can be analyzed by immobilization on membranes and hybridization with specific oligonucleotide probes, or by treatment with specific endonucleases and analysis of the products by gel electrophoresis. Labeling of the cleaved PCR products can be accomplished by incorporation of radiolabeled nucleotides, endlabeling (*e.g.*, with  $^{32}\text{P}$ -ATP), or by staining with ethidium bromide.

20           The present invention also provides methods and compositions suitable for gene therapy for individuals deficient in  $\alpha 7$  expression, production, or function. Viral vectors commonly used for *in vivo* or *ex vivo* targeting and therapy procedures are DNA-based vectors and retroviral vectors. Methods for constructing and using viral vectors are known in the art (*See, e.g.*, Miller and Rosman, BioTechn, 7:980-990, 1992). Preferably, the viral vectors are replication defective,  
25 that is, they are unable to replicate autonomously in the target cell. In general, the genome of the replication defective viral vectors which are used within the scope of the present invention lack at least one region which is necessary for the replication of the virus in the infected cell. These regions can either be eliminated (in whole or in part), be rendered non-functional by any technique known to a person skilled in the art. These techniques include the total removal,  
30 substitution (by other sequences, in particular by the inserted nucleic acid), partial deletion or addition of one or more bases to an essential (for replication) region. Such techniques may be

performed *in vitro* (i.e., on the isolated DNA) or *in situ*, using the techniques of genetic manipulation or by treatment with mutagenic agents.

Preferably, the replication defective virus retains the sequences of its genome, which are necessary for encapsidating the viral particles. DNA viral vectors include an attenuated or defective DNA virus, including, but not limited to, herpes simplex virus (HSV), papillomavirus, Epstein Barr virus (EBV), adenovirus, adeno-associated virus (AAV), and the like. Defective viruses, which entirely or almost entirely lack viral genes, are preferred, as defective virus is not infective after introduction into a cell. Use of defective viral vectors allows for administration to cells in a specific, localized area, without concern that the vector can infect other cells. Thus, a specific tissue can be specifically targeted. Examples of particular vectors include, but are not limited to, a defective herpes virus 1 (HSV1) vector (Kaplitt *et al.*, Mol. Cell. Neurosci., 2:320-330, 1991), defective herpes virus vector lacking a glycoprotein L gene (*See e.g.*, Patent Publication RD 371005 A), or other defective herpes virus vectors (*See e.g.*, International Patent Publication No. WO 94/21807; and International Patent Publication No. WO 92/05263); an attenuated adenovirus vector, such as the vector described (Stratford-Perricaudet *et al.*, J Clin Invest, 90:626-630, 1992; and La Salle *et al.*, Science, 259:988-990, 1993); and a defective adeno-associated virus vector (Samulski *et al.*, J Virol, 61:3096-3101, 1987; Samulski *et al.*, J Virol, 63:3822-3828, 1989; and Lebkowski *et al.*, Mol Cell Biol, 8:3988-3996, 1988).

Preferably, for *in vivo* administration, an appropriate immunosuppressive treatment is employed in conjunction with the viral vector (*e.g.*, adenovirus vector), to avoid immunodeactivation of the viral vector and transfected cells. For example, immunosuppressive cytokines, such as interleukin-12 (IL-12), interferon-gamma (IFN- $\gamma$ ), or anti-CD4 antibody, can be administered to block humoral or cellular immune responses to the viral vectors. In addition, it is advantageous to employ a viral vector that is engineered to express a minimal number of antigens.

In a preferred embodiment, the vector is an adenovirus vector. Adenoviruses are eukaryotic DNA viruses that can be modified to efficiently deliver a nucleic acid of the invention to a variety of cell types. Various serotypes of adenovirus exist. Of these serotypes, preference is given, within the scope of the present invention, to type 2 or type 5 human adenoviruses (Ad 2 or Ad 5), or adenoviruses of animal origin (*See*, WO94/26914). Those adenoviruses of animal origin, which can be used within the scope of the present invention, include adenoviruses of

canine, bovine, murine (*e.g.*, Mav1, described by Beard *et al.*, Virol, 75-81, 1990), ovine, porcine, avian, and simian (*e.g.*, SAV) origin. Preferably, the adenovirus of animal origin is a canine adenovirus, more preferably a CAV2 adenovirus (*e.g.*, Manhattan or A26/61 strain ATCC VR-800, for example).

5 In another embodiment the gene can be introduced in a retroviral vector (*e.g.*, as described in U.S. Patent Nos. 5,399,346, 4,650,764, 4,980,289 and 5,124,263; all of which are herein incorporated by reference; Mann *et al.*, Cell, 33:153, 1983; Markowitz *et al.*, J Virol, 62:1120, 1988; PCT/US95/14575; EP 453242; EP178220; Bernstein *et al.*, Genet Eng, 7:235, 1985; McCormick, BioTechnol, 3:689, 1985; International Patent Publication No. WO95/07358; 10 and Kuo *et al.*, Blood, 82:845, 1993). The retroviruses are integrating viruses that infect dividing cells. The retrovirus genome includes two LTRs, an encapsidation sequence and three coding regions (*gag*, *pol* and *env*). In recombinant retroviral vectors, the *gag*, *pol* and *env* genes are generally deleted, in whole or in part, and replaced with a heterologous nucleic acid sequence of interest. These vectors can be constructed from different types of retrovirus, such as, HIV, 15 MoMuLV ("murine Moloney leukaemia virus" MSV ("murine Moloney sarcoma virus"), HaSV ("Harvey sarcoma virus"); SNV ("spleen necrosis virus"); RSV ("Rous sarcoma virus") and Friend virus. Defective retroviral vectors are disclosed in WO95/02697.

In general, in order to construct recombinant retroviruses containing a nucleic acid sequence, a plasmid containing the LTRs, the encapsidation sequence and the coding sequence is 20 constructed. This construct is used to transfect a packaging cell line, which cell line is able to supply in trans the retroviral functions that are deficient in the plasmid. In general, the packaging cell lines are thus able to express the *gag*, *pol* and *env* genes. Such packaging cell lines have been described in the prior art, in particular the cell line PA317 (US Patent No. 4,861,719, herein incorporated by reference); the PsiCRIP cell line (*See*, WO90/02806), and the 25 GP+envAm-12 cell line (*See*, WO89/07150). In addition, the recombinant retroviral vectors can contain modifications within the LTRs for suppressing transcriptional activity as well as extensive encapsidation sequences which may include a part of the *gag* gene (Bender *et al.*, J Virol, 61:1639, 1987). Recombinant retroviral vectors are purified by standard techniques known to those having ordinary skill in the art.

30 Alternatively, the vector can be introduced *in vivo* by lipofection. For the past decade, there has been increasing use of liposomes for encapsulation and transfection of nucleic acids *in*

*vitro*. Synthetic cationic lipids designed to limit the difficulties and dangers encountered with liposome-mediated transfection can be used to prepare liposomes for *in vivo* transfection of a gene encoding a marker (Felgner *et al.*, Proc Natl Acad Sci USA, 84:7413-7417, 1987; Mackey, *et al.*, Proc Natl Acad Sci USA, 85:8027-8031, 1988; and Ulmer *et al.*, Science, 259:1745-1748, 5 1993). The use of cationic lipids may promote encapsulation of negatively charged nucleic acids, and also promote fusion with negatively charged cell membranes (Felgner and Ringold, Science, 337:387-388, 1989). Particularly useful lipid compounds and compositions for transfer of nucleic acids are described in International Patent Publications WO95/18863 and WO96/17823, and in U.S. Patent No. 5,459,127, herein incorporated by reference.

10 Other molecules are also useful for facilitating transfection of a nucleic acid *in vivo*, such as a cationic oligopeptide (*e.g.*, International Patent Publication WO95/21931), peptides derived from DNA binding proteins (*e.g.*, International Patent Publication WO96/25508), or a cationic polymer (*e.g.*, International Patent Publication WO95/21931).

15 It is also possible to introduce the vector *in vivo* as a naked DNA plasmid. Methods for formulating and administering naked DNA to mammalian muscle tissue are disclosed in US Patent Nos. 5,580,859 and 5,589,466, both of which are herein incorporated by reference.

DNA vectors for gene therapy can be introduced into the desired host cells by methods known in the art, including but not limited to transfection, electroporation, microinjection, transduction, cell fusion, DEAE dextran, calcium phosphate precipitation, use of a gene gun, or 20 use of a DNA vector transporter (*See e.g.*, Wu *et al.*, J Biol Chem, 267:963-967, 1992; Wu and Wu, J Biol Chem, 263:14621-14624, 1988; and Williams *et al.*, Proc Natl Acad Sci USA, 88:2726-2730, 1991). Receptor-mediated DNA delivery approaches can also be used (Curiel *et al.*, Hum Gene Ther, 3:147-154, 1992; Wu and Wu, J Biol Chem, 262:4429-4432, 1987).

25 The present invention also provides methods and compositions for the production of *in vitro* cell cultures that express wild-type or mutated human  $\alpha 7$ , as well as transgenic animals capable of expressing wild-type or mutated human  $\alpha 7$ . For example, the genomic  $\alpha 7$  clone can be expressed in mammalian cells (*e.g.*, cell lines, including but not limited to mammalian kidney cells, such as HEK). It is also contemplated that in some embodiments, the cells and animals also express other foreign genes in conjunction with the introduced  $\alpha 7$ .

30 The present invention also provides methods for producing non-human transgenic animals, comprising the steps of: a) introducing into an embryonal cell of a non-human animal a

polynucleotide sequence encoding an  $\alpha 7$  protein; b) transplanting the transgenic embryonal target cell formed thereby into a recipient female parent; and c) identifying at least one offspring containing the transgene wherein the  $\alpha 7$  mRNA is overexpressed in the tissue of the offspring.

In one preferred embodiment, the  $\alpha 7$  mRNA is human  $\alpha 7$  mRNA. In an alternative

embodiment, the polynucleotide sequence encoding an  $\alpha 7$  protein comprises a yeast artificial chromosome, while in another embodiment, the polynucleotide sequence encoding an  $\alpha 7$  is a bacterial artificial chromosome, and in yet another embodiment, the polynucleotide sequence encoding an  $\alpha 7$  protein is a P1 artificial chromosome. In a further embodiment, the non-human animal is a member of the Order Rodentia. In a preferred embodiment, the non-human animal is a mouse.

## DESCRIPTION OF THE FIGURES

Figure 1 shows the exon-intron boundary sequences of the human  $\alpha 7$  nAChR subunit gene. The 3' portion of exon 1 is disclosed as SEQ ID NO:84. Also shown are the splice acceptor sequences of: intron 1 (SEQ ID NO:85), intron 2 (SEQ ID NO:86), intron 3 (SEQ ID NO:87), intron 4 (SEQ ID NO:88), intron 5 (SEQ ID NO:89), intron 6 (SEQ ID NO:90), intron 7 (SEQ ID NO:91), intron 8 (SEQ ID NO:92) and intron 9 (SEQ ID NO:93), as well as the splice donor sequences of: intron 1 (SEQ ID NO:104), intron 2 (SEQ ID NO:106), intron 3 (SEQ ID NO:108), intron 4 (SEQ ID NO:110), intron 5 (SEQ ID NO:112), intron 6 (SEQ ID NO:114), intron 7 (SEQ ID NO:116), intron 8 (SEQ ID NO:118), and intron 9 (SEQ ID NO:120). Additionally, flanking exon sequences are shown: exon 2 (SEQ ID NO:105), exon 3 (SEQ ID NO:107), exon 4 (SEQ ID NO:109), exon 5 (SEQ ID NO:111), exon 6 (SEQ ID NO:113), exon 7 (SEQ ID NO:115), exon 8 (SEQ ID NO:117), exon 9 (SEQ ID NO:119), and exon 10 (SEQ ID NO:121).

Figure 2 shows the sequence variants identified in full-length and duplicated genomic clones.

Figure 3 provides an expression analysis of sequence variants.

Figure 4 provides the nucleotide sequence of the region 5' of the human  $\alpha 7$  nAChR subunit gene (SEQ ID NO:94).

Figure 5 shows the genomic contig of clones positive for  $\alpha 7$  nAChR gene sequences and surrounding loci.

Figure 6 provides a partial sequence of a RACE clone, with exon sequences shown in upper case and intron sequences shown in lower case: exon D (SEQ ID NO:95), exon C (SEQ ID NO:96), exon B (SEQ ID NO:97), exon A (SEQ ID NO:98), exon 5 (SEQ ID NO:99), and exon 6 (SEQ ID NO:100).

Figure 7 provides a map of the full-length  $\alpha 7$  nAChR gene. Panel A shows a physical map of the region and the structure of the full-length  $\alpha 7$  nAChR gene. Panel B shows the structure of alternatively spliced transcripts.

Figure 8 provides the DNA sequence of the human  $\alpha 7$  neuronal nicotinic receptor promoter (SEQ ID NO:101).

Figure 9 provides the DNA sequence of the alternatively spliced human  $\alpha 7$  neuronal nicotinic receptor RACE product A/C/D (SEQ ID NO:102).

Figure 10 provides the DNA sequence of the alternatively spliced human  $\alpha 7$  neuronal nicotinic receptor RACE product A/B/C/D (SEQ ID NO:103).

Figure 11 provides a physical map of the linkage region to schizophrenia on chromosome 15q13-q14. The estimated size of the region is 4 cm.

Figure 12 depicts the promoter region of the  $\alpha 7$  nicotinic acetylcholine receptor gene (*CHRNA7*). Panel A shows the fragments used to identify the core promoter region. Panel B shows the core promoter region for the *CHRNA7* gene set forth as SEQ ID NO:125. Arrows depict the locations of polymorphisms identified with the mutation screen.

Figure 13 provides the results of a functional assay of the  $\alpha 7$  core promoter variants. The activity of the normal promoter sequence was set at 100%. Symbols are indicative of the following P values: asterisk,  $P < 0.0001$ ; dagger,  $P = 0.005$ ; double dagger,  $P = 0.05$ ; and section mark,  $P = 0.03$ .

Figure 14 shows the gating of the P50 auditory evoked potential in control subjects (Panels A and B), and in a schizophrenic subject (Panel C). Tracings are shown for both the conditioning and test responses. Arrows indicate the paired auditory stimuli.

Figure 15 depicts the association between promoter variants and P50 gating in control subjects. In Panel A, mean P50 ratios are shown for control subjects without (N/N) and with one or more polymorphisms (Poly) in the  $\alpha 7$  core promoter. In Panel B, promoter variants in control subjects are shown to fit into three P50 gating ranges.

Figure 16 provides a schematic of the 15q13-q14 region containing the *CHRNA7* and *dupCHRNA7* genes in Panel A. The transcripts from both the  $\alpha 7$  containing genes are shown with their unique 5' ends in Panel B. The number of variants mapped to each exon is shown in parentheses.

Figure 17 depicts predictive patterns for 24 different mutations in the *CHRNA7* proximal promoter determined through use of a Transgenomics WAVE<sup>TM</sup> denaturing high performance liquid chromatography (DHPLC) system. The patterns are as follows: (A) wild type; (B) -194 G/C; (C) -86 C/T; (D) -46 T; (E) -46 G/T; (F) -92 G/A; (G) -143 G/A; (H) -166 C/T; (I) -178 -G; (J) -180 G/C; (K) -190 +G; (L) -191 G/A; (M) -140 +CGGG; (N) -178 +CGGGGG; (O) -241 A/G; (P) -46 G/T and -178 -G; (Q) -46 G/T and -190 +G; (R) -46 G/T and -191 G/A; (S) -46 G/T and -194 G/C; (T) -86 C/T and -194 G/C; (U) -86 C/T and -241 A/G; (V) -93 C/G and -194 C/G; (W) -178 -G and -190 +G; (X) -178 -G and -191 G/A; and (Y) -191 G/A and -194 G/C.

## DESCRIPTION OF THE INVENTION

As the methods of the prior art have been unsuccessful in providing needed information regarding the genetics of schizophrenia, an alternative strategy for genetic studies of complex diseases involving the use of a specific neurobiological characteristic of the illness (*e.g.*, as an additional phenotype more closely reflecting the effect of a single genetic alteration), was used during the development of the present invention. Such information is needed in order to provide diagnostic and treatment methods, as well as animal models for schizophrenia, as well as other psychoses. It is contemplated that such a trait is part of the inherited diathesis of the illness, which produces schizophrenia in combination with other pathogenic elements.

The present invention provides genetic information (*i.e.*, sequences, including sequence location and information on intron/exon boundaries) for the  $\alpha 7$  nicotinic receptor, as well as methods to assess the function of this receptor in normal, as well as schizophrenic individuals. The present invention also provides methods and compositions for analyzing samples from patients suspected of suffering from diverse conditions, including epilepsy (*e.g.*, juvenile myoclonic epilepsy), small cell lung carcinoma and other nicotine-dependent diseases, Prader-Willi, Angelman's syndrome, and other genetic disorders, etc. Indeed, it is not intended that the present invention be limited to schizophrenia.

The following description of the present invention is arranged as follows: (I) Inhibition of the P50 Auditory Response; (II) Neuronal Nicotinic Receptor Subunit Family; (III) Association of *CHRNA7* Promoter Variants with P50 Inhibitory Deficits; (IV) Polymorphisms in *CHRNA7* and *dupCHRNA7*; (V) Detection of *CHRNA7* and *dupCHRNA7* Alleles; (VI) Treatment and  
5 Diagnosis of Schizophrenia and Other Psychoses; and (VI) Transgenic Animals.

## **I. Inhibition of the P50 Auditory Response**

Various psychophysiological paradigms demonstrate altered brain functions in schizophrenic patients and their relatives that might reflect inherited traits (*See e.g.*, Sham *et al.*,  
10 Psychiat Genet, 4:29, 1994; De Amicis *et al.*, J Nerv Ment Dis, 174:177, 1986; Holzman *et al.*, Arch Gen Psychiat, 45:641, 1988; and Braff *et al.*, Arch Gen Psychiat, 49:206, 1992). Basic deficits in the regulation of response to sensory stimuli may underlie patients' more apparent symptoms such as hallucinations and delusions. In addition to hearing voices, patients often attend to apparently extraneous stimuli in their surroundings that normal individuals generally  
15 ignore. Such symptoms suggested that neuronal mechanisms responsible for the filtering or gating of sensory input to higher brain centers are deficient. One method developed for examining such neuronal mechanisms compares the responses to first and second of paired stimuli. The first stimulus elicits an excitatory response that also activates inhibitory mechanisms, which then diminish the excitatory response to the second stimulus. The ratio of  
20 the amplitude of the second response to the first is inversely related to the strength of inhibition.

During the development of the present invention, this method was used to study the response to auditory stimuli in schizophrenia, using an electrically positive evoked potential occurring 50 ms after an auditory stimulus (P50). Inhibition of the P50 response to a second identical stimulus (presented 500 ms after the first) has been reported to be diminished in  
25 schizophrenics (Adler *et al.*, Biol Psychiat, 17:639, 1982; Boutros and Overall, Clin Electroencephalog, 22:20, 1991; Erwin *et al.*, Biol Psychiat, 30, 430, 1991; and Judd *et al.*, Am J Psychiat, 149:488, 1992). This diminished inhibition, measured as an elevation in the ratio of P50 amplitudes, has been correlated with schizophrenics' decreased performance in a neuropsychological measure of sustained attention, as well as diminished performance in a word  
30 recognition task (Cullum *et al.*, Schizophrenia Res, 10:131, 1993; and Vinogradova *et al.*, Biol Psychiat, 39: 821, 1996).

In the development of the present invention, inhibition of the P50 response was measured in animal and related clinical investigations, to identify neurobiological mechanisms related to genes of interest, as well as a phenotype for linkage analysis to identify chromosomal areas containing genes responsible for the abnormality in schizophrenics.

5       The neurobiological inhibition of human P50 to repeated auditory stimuli was initially investigated using an auditory evoked potential recorded from the rat as a model. Both the human and rat potentials show similar decreased responses to repeated auditory stimuli (Adler *et al.*, Biol Psychiat, 21:787, 1986). Neuronal recordings identified the pyramidal neurons of the hippocampus as a major source of the rat evoked potential. These pyramidal neurons have a decremented response to repeated auditory stimuli that parallels the decrement in the evoked  
10       potential (Bickford-Wimer *et al.*, Biol Psychiat, 27:183, 1990). The decrement is lost after transection of the fimbria-fornix, a fiber tract that includes afferents to the hippocampus from cholinergic neurons in the basal forebrain (Vinogradova, in *The Hippocampus 2: Neurophysiology and Behavior*, Issacson and Pribram (eds), Plenum Press, New York, NY, pp 3-  
15       69, 1975).

      However, nicotine has been found to normalize inhibition of response in the fimbria-fornix lesioned animals (*See e.g.*, Bickford and Wear, Brain Res, 705:235, 1995). Studies with pharmacological antagonists in unlesioned animals indicate that a specific subset of nicotinic cholinergic receptors is involved in the inhibitory mechanism. The inhibition is selectively  
20       blocked by the snake toxin  $\alpha$ -bungarotoxin (Luntz-Leybman *et al.*, Brain Res, 587:130, 1992), suggesting that the receptor contains the  $\alpha 7$  nicotinic cholinergic receptor subunit, as it is the only known nicotinic receptor subunit in the mammalian brain sensitive to this toxin (Couturier *et al.*, Neuron, 5:847-856, 1990; and Schoepfer *et al.*, Neuron, 5:35, 1990). Neither scopolamine, nor mecamylamine, nor  $\alpha$ -bungarotoxin (*i.e.*, antagonists of other types of  
25       cholinergic receptors), blocked the inhibition. Receptor autoradiography using [ $^{125}$ I]-  $\alpha$ -bungarotoxin showed the most intense binding to non-pyramidal hippocampal neurons containing the inhibitory neurotransmitter  $\gamma$ -aminobutyric acid (Freedman *et al.*, J Neurosci, 13:1965, 1993). This labeling was consistent with physiological evidence that cholinergic synapses activate interneurons, which inhibit the pyramidal neuron response to the second  
30       stimulus (*See e.g.*, Miller and Freedman, Neurosci, 69:371-381, 1995; and Hershman *et al.*, Neurosci Lett, 190:133, 1995).

There are several areas of apparent concordance between these findings in rats and P50 inhibition in humans. First, P50 has been recorded from the human hippocampus (Goff *et al.*, Prog Clin Neurophysiol, 7:126, 1980; and Makela *et al.*, Electroencephalogr Clin Neurophysiol, 92:414, 1994), and human hippocampal neurons have rapidly decreasing responses to auditory stimuli, similar to those observed with rat hippocampal neurons (Wilson *et al.*, Exp Neurol, 84:74, 1984). Second, nicotine in high doses transiently normalizes the abnormality in P50 inhibition in schizophrenics and in their relatives, much as it normalizes inhibition in rats after fimbria-fornix lesions (Bickford and Wear, *supra*, 1995; Adler *et al.*, Biol Psychiatry, 32:607, 1992; and Adler *et al.*, Am J Psychiat, 150:1856, 1993). However, the effect of nicotine on P50 inhibition in relatives of schizophrenics is not blocked by mecamylamine, which blocks all known nicotinic receptors in human brain, except the  $\alpha 7$  nicotinic receptor (Freedman *et al.*, Harvard Rev Psychiat, 2:179, 1994). *In situ* hybridization has shown that  $\alpha 7$  nicotinic receptor mRNA is expressed in human hippocampal neurons (Freedman *et al.*, Harvard Rev Psychiat, 2:179, 1994). Some of the non-pyramidal neurons of the human hippocampus were intensely labeled by  $\alpha$ -bungarotoxin, as was also observed with rats.

A preliminary study showed that  $\alpha$ -bungarotoxin labeling was decreased in post mortem hippocampus from eight schizophrenics (Freedman *et al.*, Biol Psychiat, 38:22, 1995). In addition, schizophrenic patients are particularly heavy tobacco smokers, even when compared to other psychiatric patients (deLeon *et al.*, Am J Psychiat, 152:453, 1995; and Hamera *et al.*, J Nerv Mental Dis, 183:559, 1995). This heavy nicotine use may reflect an attempt at self-medication of an endogenous neuronal deficit (Goff *et al.*, Am J Psychiat, 149:1189, 1992). However, nicotine's efficacy as an anti-psychotic is limited, due to rapid desensitization and cardiovascular toxicity.

In parallel with these biological studies in human and animals, the P50 evoked potential abnormality was also investigated as a phenotype for genetic linkage analysis. A genome-wide scan was initiated, independent of any candidate gene hypothesis, in nine multiplex schizophrenic pedigrees, which were also phenotyped with P50 recordings. The deficit in inhibition of the P50 response in these and other schizophrenic families is generally found in one of the parents and half the siblings, including the schizophrenic probands (Siegel *et al.*, Arch Gen Psychiat, 41:607-612, 1984). Although elevated P50 ratios are significantly associated with the apparent genetic risk for schizophrenia, many individuals in the pedigrees who have the deficit

are clinically unaffected (Waldo *et al.*, Psychiat Res, 39:257, 1991). In this respect, the distribution of the trait resembles several other neurobiological abnormalities in schizophrenics and their relatives, such as deficits in smooth pursuit eye movements and reaction time (De Amicis *et al.*, J Nerv Ment Dis, 174:177, 1986; and Holzman *et al.* Arch Gen Psychiat, 45:641, 1988). These traits may represent alternative expressions of a latent trait or endophenotype, which, in combination with other pathogenic elements, gives rise to schizophrenia.

During the development of the present invention, preliminary linkage analyses between the P50 ratio abnormality and 318 restriction fragment length polymorphism and tandem repeat DNA markers in the nine kindreds were conducted. DNA markers mapping to four chromosomal regions, one of which was 15q14, revealed small positive lod scores (maximum logarithm of the odds) assuming autosomal dominant transmission. Subsequently, the  $\alpha 7$  nicotinic receptor gene was localized to the 15q14 region (Orr-Urtreger *et al.*, Genomics, 26:399, 1995; and Spitzer *et al.*, Arch Gen Psychiat, 35:773, 1978). As converging evidence from neurobiological investigations implicated  $\alpha 7$  receptor function in abnormal P50 inhibition, and the preliminary linkage study provided suggestive evidence for heritability of the trait near the chromosomal location of the  $\alpha 7$  receptor gene, additional experiments, using informative markers at the  $\alpha 7$  receptor gene locus were undertaken. Two new DNA polymorphic loci were isolated, namely D15S1360 from a yeast artificial chromosome (YAC) containing the  $\alpha 7$  nicotinic receptor gene, and L76630 from an  $\alpha 7$ -containing clone in a genomic phage library. These markers were used with over 500 highly polymorphic markers in a 10 centiMorgan resolution genome-wide scan of the nine pedigrees. The results demonstrate a highly significant linkage between D15S1360 and the abnormality in P50 suppression.

Indeed, the data obtained during the development of the present invention strongly suggest that the P50 auditory sensory deficit in schizophrenia is genetically linked to the locus of the  $\alpha 7$  nicotinic receptor gene on chromosome 15q14. Thus, the present invention provides a heretofore unknown linkage between nicotinic receptors and schizophrenia. The significant linkage obtained with the P50 ratio phenotype supports the value of this strategy. This provides methods for assessing the effects of therapy to correct abnormalities in  $\alpha 7$  structure and/or function, as well as providing methods for developing and identifying drugs suitable for use in treating such abnormalities.

Although an understanding of the mechanism is not necessary in order to use the present invention, it has been suggested that the clinical illness may be less penetrant, because multiple genetic and non-genetic factors are required to produce clinical illness, whereas a specific biological defect may occur as the result of a single gene effect. Thus, some gene carriers would be expected to have abnormal P50 ratio, the more penetrant phenotype, but not schizophrenia, which is less penetrant. The lower lod scores observed during the development of the present invention with schizophrenia as a phenotype support that position; several kindreds had higher lod scores for P50 ratio than for schizophrenia because there were many family members with abnormal P50 ratios who did not have schizophrenia.

The possibility that the chromosome 15q13-14 region is involved in psychotic illness has also been investigated in relationship to other diseases. For example, psychoses resembling schizophrenia have been observed in Prader-Willi syndrome, a mental retardation linked to deletions and abnormal DNA imprinting in the 15q11-13 region (Clarke, *Brit J Psychiat*, 163:680, 1993). The imprinting abnormality affects the expression of many genes in this region. Several families in Sephardic and other populations have co-existent schizophrenia and Marfan's syndrome (*i.e.*, a disease linked to dominant mutations in the fibrillin gene at 15q21; Sirota *et al.*, *Br J Psychiat*, 157: 433, 1990; and Melissari *et al.*, *Pathologica*, 87:78, 1995). The co-segregation of the two illnesses may be based on their chromosomal proximity. Psychosis was also observed in a large French-Canadian kindred that has a recessive demyelination disease, linked to markers at 15q14 (Casaubon *et al.*, *Am J Hum Genet*, 58:28, 1996). In addition, an Italian kindred contains two cousins with psychotic illness and a partial trisomy of chromosome 15, derived independently from abnormal meioses involving a balanced familial translocation with a 15q13 breakpoint, that was present in each of their mothers. It was suggested that the new trisomies may have caused the *de novo* appearance of illness (Calzolari *et al.*, *Am J Med Genet*, 67:154, 1996). The present invention provides the means to determine to what extent the appearance of psychoses in these families with other genetic abnormalities at 15q13-14 involves the  $\alpha 7$  gene.

In addition to providing means to assess the risk for development of schizophrenia, the present invention also provides new data about the identity of neuronal abnormalities involved in its pathophysiology, as well as the means to develop treatment methods and compounds, diagnostic methods and reagents, and models (*e.g.*, cell lines and transgenic animals) of these

neuronal abnormalities. These results are consistent with clinical and neurobiological evidence for the involvement of the  $\alpha 7$  nicotinic receptor gene in sensory gating deficits in schizophrenia.

The present invention also provides the means to determine the role the  $\alpha 7$  receptor in the sensory processing defects and other abnormalities in schizophrenia. The finding of a significant linkage supporting the role of the  $\alpha 7$  nicotinic receptor in the pathophysiology of sensory and attentional disturbance in schizophrenia, is unique. Many neurotransmitter systems have been hypothesized to be at least partly responsible for schizophrenia, but direct biological assessment of a specific neuronal receptor function in human subjects is generally not feasible because of the brain's complexity and inaccessibility. The present invention provides compositions and methods to overcome these drawbacks. Genetic investigations, including linkage studies, have represented the critical test of the involvement of a particular mechanism in schizophrenia. The present invention provides methods and compositions to complement and/or replace such tests for schizophrenia. Indeed, linkage at the  $\alpha 7$  nicotinic receptor locus thus supports the neurobiological evidence that this gene plays a role in a pathophysiological aspect of schizophrenia, a role that prior to the present invention, had not been previously considered nor described, despite schizophrenics' well-known heavy dependence on nicotine.

## **II. Neuronal Nicotinic Receptor Subunit Family**

As discussed above, during the development of the present invention, the  $\alpha 7$  nicotinic receptor was associated with pathophysiological aspect(s) of schizophrenia. This receptor is a member of the neuronal nicotinic receptor subunit gene family, which is expressed in mammalian brain as pentameric, ligand-gated ion channels (Patrick *et al.*, Ann NY Acad Sci, 505:194, 1987; Cooper *et al.*, Nature, 350:235, 1991; and Lindstrom *et al.*, Ann NY Acad Sci, 757:100, 1996). In the muscle, five different types of subunits constitute the holoreceptor, but in brain only two types of subunits, designated as " $\alpha$ " and " $\beta$ ," have been found (Galzi *et al.*, Ann Rev Pharmacol, 31:37, 1991; and Lukas and Bencherif, Int Rev Neurobiol, 34:25, 1992).

Neuronal receptors can be functionally differentiated into two principal classes, which differ in their affinity for nicotine and the snake toxin,  $\alpha$ -bungarotoxin (Marks and Collins, Mol Pharmacol, 22:554, 1982; Wonnacott, J Neurochem, 47:1706, 1986; Marks *et al.*, Mol Pharmacol, 30:427, 1986; and Amar *et al.*, FEBS, 327:284, 1993). Receptors that bind nicotine with high affinity contain  $\alpha 2$ - $\alpha 6$  as ligand binding subunits and require an association with  $\beta$

subunits for functional expression (Goldman *et al.*, Cell, 48:965, 1987; Deneris *et al.*, Clin Chem, 35:731, 1989; and Wada *et al.*, J Compar Neurol, 284:314, 1989). A second class of receptors ( $\alpha 7$ - $\alpha 9$ ) bind nicotine with low affinity, have a high affinity for  $\alpha$ -bungarotoxin, and function as homomeric ion channels in *in vitro* expression systems (Marks *et al.*, *supra*, 1986; Wonnacott, *supra*, 1986; Alkondon and Albuquerque, J Pharm Ex Ther, 265:1455, 1993; Amar *et al.*, FEBS, 327:284, 1993; and Zhang *et al.*, Neuron, 12:167, 1994). The  $\alpha 7$  receptor is the only  $\alpha$ -bungarotoxin-binding receptor identified in mammalian brain. The  $\alpha 8$  receptor appears to be only expressed in chick (Schoepfer *et al.*, Neuron, 5:35, 1990), and the  $\alpha 9$  receptor has limited expression in cochlear hair cells and pituitary (Elgoyhen *et al.*, Cell, 79:705, 1994). In addition, a cDNA clone of the human  $\alpha 7$  was isolated from a human brain library (GenBank Accession No. U40583).

Expression and function of a specific subset of the nicotinic receptor family, the  $\alpha 7$  receptor, has recently been implicated in a neuronal pathway controlling the filtering or gating of auditory stimuli in both human and rat brain (Adler *et al.*, Biol Psych, 32:607, 1992; Adler *et al.*, Am J Psychol, 150:1856, 1993; Freedman *et al.*, Harvard Rev Psychiat, 2:179, 1994; and Leonard *et al.*, *supra* 1996). This sensory processing mechanism is aberrant in a majority of subjects with schizophrenia (Freedman *et al.*, Schiz Res, 4:233-243, 1991). Pharmacological studies in both humans and rats suggest that the deficit in humans can be normalized by nicotine (Adler *et al.*, *supra*, 1992; and Adler *et al.*, *supra* 1993) and reproduced in a rodent model by antagonists of the low affinity  $\alpha 7$  nicotinic receptor but not by high affinity antagonists (Luntz-Leybman *et al.*, Brain Res, 587:130, 1992; and Rollins *et al.*, Soc Neurosci Abst, 22:1272, 1996). Expression of  $\alpha$ -bungarotoxin binding receptors is decreased in schizophrenic hippocampi by approximately 40% (Freedman *et al.*, Biol. Psychiat, 38:22, 1995).

During the development of the present invention, the locus D15S1360, a polymorphic marker <120 kb from the full-length  $\alpha 7$  nicotinic receptor gene at 15q14, was genetically linked to this auditory gating deficit in schizophrenic pedigrees. However, it is contemplated that other genes mapping to the 15q14 region are potential alternative or additional genetic candidates to  $\alpha 7$  for pathogenic features of schizophrenia.

Also during the development of the present invention, expression of the  $\alpha 7$  nicotinic receptor was examined in human postmortem brain. This gene was widely expressed in most

nuclei, albeit at low levels. Regions of highest expression included those involved in processing of sensory information, such as the hippocampus, lateral and medial geniculates, and the reticular nucleus of the thalamus.

The present invention further provides the physical mapping of a full-length human  
5 genomic clone for the  $\alpha 7$  receptor subunit and sequencing of a putative promoter region. The gene and promoter structure are similar to that of the chick  $\alpha 7$ . Additionally, a partial  $\alpha 7$  gene duplication including exons 5-10 and intervening intronic sequence, which lies <1 Mb from the full-length gene has been identified. In addition, four novel exons at the 5' end of the duplicated  $\alpha 7$  sequences were sequenced and intron/exon junctions identified. The duplicated  $\alpha 7$  sequences  
10 were found to be expressed as alternatively spliced transcripts containing some or all of these novel exons.

The present invention also provides the structural organization of the human  $\alpha 7$  neuronal nicotinic acetylcholine receptor gene, and presents data indicating a partial gene duplication. Large insert genomic clones were isolated from YAC, BAC and PAC libraries. There are 10  
15 exons in the gene; the splice junctions are consistent with consensus splice sites (Green, *Ann Rev Cell Biol*, 7:559, 1991; Lamond, *Bioessays*, 15:595, 1993) and have an identical location to those in the chick  $\alpha 7$  gene (Matter-Sadzinski *et al.*, *EMBO J*, 11:4529, 1992), the only species for which genomic  $\alpha 7$  gene sequence has been previously published.

The promoter region of the gene was found to be 77% G/C, and contains no TATA box.  
20 It thus fits a growing group of eukaryotic promoters, which demonstrate multiple transcription start sites (Maue *et al.*, *Neuron*, 4:223, 1990; and Sauerwald *et al.*, *J Biol Chem*, 265:14932, 1990). The nucleotide sequence, between the human and chick promoter regions, was poorly conserved. However, there are consensus transcription factor binding sites located in similar positions in the two promoters (Matter-Sadzinski *et al.*, *supra*, 1992). These include SP-1 and  
25 AP-2 binding sites. SP-1 and AP-2 consensus motifs are frequently found in other ligand-gated ion channel genes (*See e.g.*, Bessis *et al.*, *Nucl Acids Res*, 21:2185, 1993), and may contribute to neuronal specificity.

A cyclic AMP response element (CREB) binding site motif was also identified in the human promoter, but is not found in the chick gene. The presence of this CREB site in the  
30 human promoter is interesting since the mammalian  $\alpha 7$  gene is known to be down-regulated by corticosterone (Pauly *et al.*, "Glucocorticoid Regulation of Sensitivity to Nicotine," *in The*

*Biology of Nicotine: Current Research Issues*, Raven Press, New York, NY, pp. 121-139, 1992), which affects expression of the CREB-binding protein. Thus, it is contemplated that corticosterone and other glucocorticoids affect the  $\alpha 7$  gene in some embodiments of the present invention.

5 In addition, alternative splicing of the full-length  $\alpha 7$  gene was detected during the development of the present invention. Six different splice variants were identified through sequencing full-length transcripts. However, only one, missing exon 3, did not interrupt the frame of translation.

10 Several important motifs, which affect correct splicing of heterogeneous RNA, were identified during the development of the present invention. For example, there are two Chi(X) sequences (consensus: CCTGGTGG) known to enhance splicing, present in the human  $\alpha 7$  gene of the present invention; there is one in intron 4 and one in the 3'-UT of the cDNA. Another group of splice enhancers with sequence (T)GCATG(A), have been localized as well. There are seven motifs of this enhancer class in sequence identified for intron 2 (approximately >25 kb in  
15 size). An additional enhancer of this type has been found in the large intron 4. It is contemplated that additional splicing motifs are localized in the human  $\alpha 7$  sequence.

Exons 5-10 of the  $\alpha 7$  nicotinic receptor gene were found to be duplicated in the human genome. The duplicated sequences lie within 1 Mb and are centromeric to the full-length  $\alpha 7$  gene on chromosome 15. The evidence for the duplication includes mapping of the duplicated  
20 sequences to a different site on a YAC contig of the region. Additionally, heterozygous polymorphic sequences at exonic sites and at the L76630 locus, located 1.4 kb beyond the 3' end of the coding region, were detected in both a somatic cell chromosome 15 hybrid and in a single YAC (969b11) containing both the full-length gene and duplicated  $\alpha 7$  exons. The apparent arrangement of the duplication is head to tail in relation to the full-length gene.

25 Further complexity for the  $\alpha 7$  gene structure was introduced when it was determined that some of the RACE clones isolated during cloning of a human  $\alpha 7$  cDNA contained only exons 5-10, and additional non- $\alpha 7$  sequences 5' of exon 5. These sequences were identical to sequences found in several EST clones that were located by homology screening with  $\alpha 7$  cDNA sequence. The EST clones also contained only exons 5-10 of the  $\alpha 7$  gene, with the previously unreported  
30 sequences again 5' of exon 5. PCR products from genomic DNA and from YACs 948a10 and

953g6 revealed that these non- $\alpha 7$  sequences were present in genomic clones containing either the full-length gene or the duplicated  $\alpha 7$  sequences, and four novel exons were defined. It is contemplated that these sequences are arranged as alternatively spliced exons, as the positions of the consensus splice junctions between them correspond to the spliced products seen in the RACE and EST clones. These new exons were designated as "3'- $\alpha 7A$ ," " $\alpha 7B$ ," " $\alpha 7C$ " and " $\alpha 7D-5'$ ." The RACE products were variable in their inclusion of exon B, similar to the EST clones.

Partial gene duplication has been implicated in human disease (*See e.g.*, Hu and Worton, Hum Mutat, 1:3, 1992; Lehrman *et al.*, Cell, 48:827, 1987; and Den-Dunnen, *et al.*, Am J Hum Genet, 45:835, 1989). Thus, it is contemplated that although transcription of mRNAs containing the novel exons was found to occur at levels similar to those of the full-length coding region, the novel exons may be expressed only from the duplicated  $\alpha 7$  sequences. There is also evidence for novel exons in another gene on chromosome 15, the small nuclear riboprotein-N (SNRPN); these exons at both the 5'- and 3'-ends of the SNRPN gene are also transcribed as alternative mRNAs. In fact, *dupCHRNA7* is missing in some people, and the deletion of the duplicated gene is more commonly observed in schizophrenics than in control subjects.

It is contemplated that the human alternative transcripts containing the novel exons  $\alpha 7D$ ,  $\alpha 7C$ ,  $\alpha 7B$ , and  $\alpha 7A$ , might be translated. These alternatives lack the  $\alpha 7$  signal peptide and disulfide bridge, which have been shown to be necessary for assembly of the homologous alpha subunit in muscle (Blount and Merlie, J Cell Biol, 111:2613, 1990). However, an alternatively spliced transcript of the muscle alpha, containing an additional exon, is expressed at equal levels to the correctly spliced isoform. It is also translated, but not assembled and is localized to the endoplasmic reticulum (Beeson *et al.*, EMBO J, 9:2101, 1990; and Newland *et al.*, J Physiol, 489:767, 1995). It is contemplated that a similar localization occurs for the human  $\alpha 7$  alternative transcripts, containing the novel exons, if translated. However, it is not intended that the present invention be limited to any particular localization of these alternative transcripts.

Antibodies to the cytoplasmic loop of the chick  $\alpha 7$ , between membrane spanning regions III and IV, have been shown to detect  $\alpha 7$  protein in pyramidal cells of rat hippocampus (Dominguez del Toro *et al.*, J Comp Neurol, 349:325, 1994). However, during the development of the present invention, no  $\alpha$ -bungarotoxin binding (*i.e.*, indicative of a functional receptor),

was observed on the plasma membranes of these cells. Since protein, translated from alternative  $\alpha 7$  mRNAs containing D-C-B-A-5-10, would have the epitope used as antigen for cytoplasmic loop antibodies, it is possible that sequestered, but dysfunctional  $\alpha 7$  protein internally localized would be detected as well as cell surface protein. The abundance of the D-C-B-A-5-10  
5 alternative transcripts, thus, raises the possibility that they are regulatory for functional expression of  $\alpha 7$  nicotinic receptors.

Although the mechanism responsible for the gene duplication is unclear, and an understanding of the mechanism is not necessary in order to use the present invention, two *alu* repeats were found in the genomic clones. One is located in intron 4, 500 bp upstream of exon 5.  
10 The second is located in the 3'-end of the gene outside of the poly-adenylation site. *Alu* repeats are known to have several possible functions, including as either positive or negative enhancers of transcription. In addition, they have also been shown to mediate duplication or deletion of DNA sequences (Schmid, Prog Nucl Acid Res, 53:283, 1996; and Lehrman *et al.*, Cell, 48:827, 1987).

15 It is clear that the duplicated and expressed sequences involving the human  $\alpha 7$  nicotinic receptor gene of the present invention provide the methods and compositions needed for mutation screening in disease. The present invention also provides methods and compositions for treatment (including, but not limited to gene therapy) of deficits in  $\alpha 7$  expression and/or function.

20 The present invention provides methods and compositions needed to determine the control of  $\alpha 7$  expression, through the use of the DNA sequences in its promoter region, as well as DNA sequences located at its intron/exon boundaries, and DNA sequences present elsewhere in its introns. In addition, the present invention provides the locations and sequences of newly identified duplicated and additional exons. It is contemplated that these sequences may be  
25 involved in pathogenic mutational events. Although the coding sequence of  $\alpha 7$  shares some similarities between various animals (*e.g.*, chickens, rodents, and humans), the genomic structure provided in the present invention in the promoter and introns is unique to humans, and could not have been predicted based on the knowledge of the genome structure of any other species.

Furthermore, the coding region alone cannot be used for genetic screening of individuals  
30 to identify mutations, because the appropriate primers (*e.g.*, for PCR) are needed from introns positioned outside of the coding region. In addition, the genomic sequence is necessary for the

production of cell lines and transgenic animals (*i.e.*, for models useful for the development of therapeutic targets in drug discovery). The present invention provides the needed genomic sequences and primers for genetic screening methods and drug discovery.

### III. Association of *CHRNA7* Promoter Variants with P50 Inhibitory Deficits

Schizophrenia is a complex disorder, in which heterogeneity, reduced penetrance, and environmental factors have made identification of genetic defects difficult. The work of many investigators has resulted in the discovery and replication of eight principal linkage regions in the human genome. These include linkages at chromosome 1q21-q22 (Brzustowicz *et al.*, Science, 288:678-682, 2000), chromosome 6p22-p24 (Straub *et al.*, Mol Psychiatry, 1:89-92, 1996), chromosome 6q21-q22 (Cao *et al.*, Genomics, 43:1-8, 1997), chromosome 8p21-p22 (Blouin *et al.*, Nat Genet, 20:70-73, 1998), chromosome 10p11-p15 (Faraone *et al.*, Am J Med Genet, 81:290-295, 1998), chromosome 13q14-q32 (Blouin *et al.*, *supra*, 1998), chromosome 15q13-ql5 (Riley *et al.*, Am J Med Genet, 96:196-201, 2000; Stober *et al.*, Am J Med Genet, 67:1201-1207, 2000, Stassen *et al.*, Am J Med Genet, 96:173-177, 2000; Freedman *et al.*, Am J Med Genet, 105:20-22, 2001; Liu *et al.*, Am J Med Genet, 105:658-661, 2001; Xu *et al.*, Am J Med Genet, 105:669-674, 2001; Tsuang *et al.*, Am J Med Genet, 105:662-668, 2001; and Gejman *et al.*, Am J Med Genet, 105:789-793, 2001), and chromosome 22q11-q13 (Pulver *et al.*, Am J Med Genet, 54:36-43, 1994). Additional linkages on six other chromosomes may be contributory in some populations (Baron, Am J Med Genet, 68:299-312, 2001). In general, linkage in any given cohort is found in only a subset of the total number of families examined, indicating that abnormalities in different gene sets may result in the same illness. Identification of pathogenic mutations in candidate genes that lie in the major linkage regions is necessary for a rigorous understanding of how several genes interact in the development of schizophrenia. As is described in more detail in Examples 11-16, the inventors provide evidence that functional polymorphisms in the promoter region of the  $\alpha 7$  neuronal nicotinic acetylcholine receptor subunit gene (*CHRNA7* or  $\alpha 7$ ), a candidate gene in the 15q13-ql4 linkage region, were more frequently found in schizophrenic patients and were associated with a sensory deficit found in this common mental illness.

The *CHRNA7* gene cluster maps to a region of replicated linkage in schizophrenia on chromosome 15q13-ql4 (*See*, Figure 11). D15S1360, a polymorphic marker in intron 2 of the

*CHRNA7* gene, is genetically linked to a sensory deficit trait in the disease, namely a failure to inhibit the response to repeated auditory stimuli in the immediate environment (lod = 5.3,  $\theta$  = 0.0,  $P < 0.001$  as described Freedman *et al.*, Proc Natl Acad Sci USA, 94:587-592, 1997).

Linkage to schizophrenia was also positive in this study of nine families, although not as

5 significant (lod = 1.33). Additional evidence for linkage of this locus to schizophrenia as the phenotype was found in pedigrees from the National Institute of Mental Health (NIMH) Schizophrenia Genetics Initiative (Freedman *et al.*, Am J Med Genet, 105:20-22, 2001; and Leonard *et al.*, Am J Med Genet, 81:308-312, 1998). A sibpair analysis showed that a significant proportion of D15S1360 alleles were shared identical-by-descent in the schizophren-

10 ics (0.58;  $P < 0.0024$ ). In a transmission disequilibrium study of schizophrenia, significant genotype-wise disequilibrium ( $P < 0.007$ ) was found at D15S165, a polymorphic simple sequence marker localized within one megabase of the  $\alpha 7$  nicotinic receptor gene at 15q13-q14 (Freedman *et al.*, Am J Med Genet, 105:20-22, 2001). Recently a full genomic linkage analysis was

15 completed of the NIMH Schizophrenia Initiative pedigrees, for which the genotyping was available from Millenium Pharmaceuticals (Cambridge, MA). A parametric genetic analysis and an autosomal codominant model was used, with a diagnosis of schizophrenia and schizoaffective disorder, depressed type, as the affected phenotype. One genetic linkage was found significant by genome-wide criteria (multipoint lod score, 3.94;  $P = 0.00005$ ), to the locus on 15q13-q14, within one cM of the previous finding for linkage to the locus of the  $\alpha 7$  nicotinic receptor gene

20 (Freedman and Leonard, Schizophr Res, 49:70, 2001; and Freedman *et al.*, Am J Med Genet, 105:794-800, 2001). Several different groups have independently replicated this finding by using nonparametric methods in the NIMH sample (Kaufmann *et al.*, Am J Med Genet, 81:282-289, 1998), and in other samples. The same region has been linked to juvenile myoclonic epilepsy (Elmslie *et al.*, Hum Mol Genet, 6:1329-1334, 1997) and more recently to bipolar disorder

25 (Edenberg *et al.*, Am J Med Gen, 74:238-246, 1997; and Turecki *et al.*, Mol Psychiatry, 6:570-578, 2001), indicating that the locus may contain defects in a gene or genes common to several neuronal disorders.

Biological and pharmacologic evidence also supports the *CHRNA7* gene as a candidate gene for schizophrenia (Adler *et al.*, Schizophr Bull, 24:189-202, 1998; and Leonard *et al.*, "The role of nicotine and nicotinic receptors in psychopathology," in Arneric and Brioni (eds.)

30 Neuronal Nicotinic Receptors: Pharmacology and Therapeutic Opportunities. New York,

NY:Wiley-Liss Inc., pp305-320, 1999). Nicotine normalizes a sensory gating abnormality, the P50 inhibitory deficit, found in most patients with schizophrenia and in 50% of their first-degree relatives (Adler *et al.*, Biol Psychiatry, 32:607-616, 1992; Adler *et al.*, Am J Psychiatry, 150:1856-1861, 1993; and Freedman *et al.*, Schizophr Res, 4:233-243, 1991). This trait, which involves inhibition of the response to repeated stimuli presented through the auditory system to the brain, can be measured by means of auditory evoked potentials in a paired pulse paradigm. Electrodes on the scalp record waves with a 50 millisecond latency (P50) following paired auditory stimuli delivered 0.5 second apart (Freedman *et al.*, Schizophr Res, 4:233-243, 1991; and Baker *et al.*, Biol Psychiatry, 22:603-617, 1987). In a normal response, the subject decreases the amplitude of the second response (test response), compared with the response to the first stimulus (conditioning response), through the action of an inhibitory neuronal pathway. The results are reported as the P50 test-conditioning (T/C) ratio. More than 85% of schizophrenic patients have abnormally increased P50 ratios, where the test response is greater than expected in the normal population (Adler *et al.*, Biol Psychiatry, 17:639-654, 1982; Clementz *et al.*, Schizophr Res, 30:71-80, 1998; Yee *et al.*, J Abnorm Psychol, 107:691-698, 1998; Erwin *et al.*, Schizophr Res, 33:157-167, 1998; and Patterson *et al.*, Arch Gen Psychiatry, 57:57-64, 2000). This P50 inhibitory deficit is inherited in families of schizophrenic patients in an apparently autosomal dominant pattern (Freedman *et al.*, Proc Natl Acad Sci USA, 94:587-592, 1997; Freedman *et al.*, Somat Cell Mol Genet, 13:479-484, 1987; and Clementz *et al.*, Am J Psychiatry, 155:1691-1694, 1998). Thus, half of family members have aberrant gating of the P50 auditory evoked potential, whether or not they have the disease. The increased incidence in schizophrenic patients and their families indicates that the P50 deficit represents a trait that predisposes to schizophrenia. The deficit is also present, but at much lower levels, in the general population, in subjects with no familial history of schizophrenia (Waldo *et al.*, Biol Psychiatry, 47:231-239, 2000). The P50 inhibitory deficit, as previously discussed, is also genetically linked to 15q13-ql4 (Freedman *et al.*, Proc Natl Acad Sci USA, 94:587-592, 1997; and Coon *et al.*, Biol Psychiatry, 34:277-289, 1993).

It is contemplated that the deficit in P50 inhibition, reflects decreased activity or expression of the CHRNA7 receptor. Pharmacologic antagonists of the CHRNA7 receptor reproduce the inhibitory deficit in several animal models (Luntz-Leybman *et al.*, Brain Res, 587:130-136, 1992; and Rollins *et al.*, Soc Neurosci Abstr, 19:837, 1993). The DBA/2j mouse

strain has 50% lower levels of *CHRNA7* than most other inbred strains, it does not show inhibition of its auditory evoked response to repeated stimuli, and the inhibition is normalized by both nicotine and a specific agonist of the  $\alpha 7$  receptor, 2,4-dimethoxybenzylidene anabaseine (Stevens *et al.*, Neuropsychopharm, 15:152-162, 1996; and Stevens *et al.*, Psychopharmacology, 136:320-327, 1998). The inventors have previously found that the expression of the *CHRNA7* gene was also decreased by approximately 50% in human postmortem hippocampus isolated from schizophrenic subjects, compared with matched control subjects (Freedman *et al.*, Biol Psychiatry, 38:22-33, 1995). Expression of the *CHRNA7* gene was also decreased in different brain regions, including the reticular thalamic nucleus and frontal cortex, in schizophrenic subjects (Court *et al.*, J Neurochem, 73:15980-1597, 1999; and Guan *et al.*, Neuroreport, 10:1779-1782, 1999).

As described in more detail in Examples 1- 10, a genomic clone for the human *CHRNA7* subunit was isolated from a yeast artificial chromosome (YAC) library. Mapping of the gene showed that exons 5 to 10 of the *CHRNA7* gene were duplicated as part of a large DNA cassette. The duplication was inserted approximately one Mb proximal to the full-length  $\alpha 7$  gene and directly 3' of five novel exons (D'-D-C-B-A). The duplicated exons 5 to 10 are expressed with the novel exons D'-A (*dupCHRNA7*) as messenger RNA in both human brain and peripheral tissues. Interestingly, *dupCHRNA7* was homozygotically missing in five (4.2%) of 118 schizophrenic patients, but not in 59 control subjects. Mutation screening of the amino acid coding region for the full-length *CHRNA7* and *dupCHRNA7* genes, and a core promoter region for the full-length gene, has been completed during development of the present invention. Although multiple polymorphisms were found in the coding region, almost all were synonymous.

In addition, a core promoter region for the full-length *CHRNA7* gene was isolated that supports efficient transcription of the reporter gene, luciferase. This 231 base pair fragment contains consensus binding sites for a number of transcription factors, including stimulating protein Sp1, activator protein AP-4, and a corticosteroid-responsive element, SRE as determined by using MatInspector (Quandt *et al.*, Nucleic Acids Res, 23:4878-4884, 1995). The regions near the Sp1 binding sites contain several G/C-rich areas, which are contemplated to be binding sites for other transcription factors, such as Egr1. The location and spacing of these sites with respect to the start of exon 1 are conserved in the bovine  $\alpha 7$  gene, where they have been shown

to regulate transcription (Carrasco-Serrano *et al.*, J Biol Chem, 273:20021-20028, 1998). During development of the present invention, mutation screening of this fragment in human DNA samples from control and schizophrenic subjects, showed a large cluster of polymorphisms, many lying in these putative transcription factor binding sites.

Although schizophrenia has a large genetic component, it is thought to be oligogenic (Freedman *et al.*, Am J Med Genet, 105:794-800, 2001; and Gershon, Biol Psychiatry, 47:240-244, 2000). Heterogeneity in the inheritance of predisposing traits further complicates the orderly process of gene identification. At present, there are 14 chromosomes on which genetic linkage to schizophrenia has been identified or is suggested (Baron, Am J Hum Genet, 68:299-312, 2001). Many of these regions are contemplated to contain a gene variant contributing to the disease in the linked populations, indicating that many genes may interact in the disorder, but that not all the gene variants at these loci may be present in a single individual. Furthermore, the actual polymorphism present in any given gene is contemplated to result in differences in gene expression between subjects, which can also be affected by other genes and environmental factors. Some variants are contemplated to manifest in early development and some during puberty or postpuberty, when schizophrenia is usually first diagnosed. Additionally, some gene variants are contemplated to compensate for others, or to actually have a beneficial effect.

Three principal issues contribute to a discussion of the present invention. First, the study of a candidate gene for an endophenotype in schizophrenia, rather than the multigenic disease itself, has permitted the identification herein of a single gene defect. Endophenotypic traits found in complex disorders have been examined in attempts to simplify the biology and genetics of schizophrenia (Venables, "Input dysfunction in schizophrenia," in Maher (ed.) Progress in Experimental Personality Research, New York, NY:Academic Press, pp.1-47, 1964; and Freedman *et al.*, Biol Psychiatry, 45:551-558, 1999). Examples of such traits are inhibitory gating of the P50 auditory evoked response (Freedman *et al.*, Biol Psychiatry, 45:551-558, 1999; and Freedman *et al.*, Somat Cell Mol Genet, 13:479-484, 1987), and smooth-pursuit eye tracking (Holzman, Int Rev Neurobiol, 27:179-205, 1985; and Holzman *et al.*, Arch Gen Psychiatry, 45:641-647, 1988), both of which are found in the general population at lower levels than in the disease. In control subjects with no history of psychosis, variants in only one or a few different genes may be required to produce a specific abnormal phenotype or trait. In a disease such as schizophrenia, interdependence of multiple neurotransmitters in a single brain pathway and the

presence of multiple gene defects may worsen performance in a given quantitative trait.

However, even in schizophrenia, only a subset of the genes involved in the full clinical diagnosis is contemplated to be associated with a specific endophenotype.

Second, the nicotinic acetylcholine receptor subunit gene, *CHRNA7*, was implicated as a candidate gene in the 15q13-q15 linkage region for schizophrenia by genetic and biological data, supporting its role in sensory processing deficits in the disease (Leonard *et al.*, Eur J Pharmacol, 393:237-242, 2000; and Leonard *et al.*, Restor Neurol Neurosci, 12:195-201, 1998). Expression of the *CHRNA7* gene is decreased in postmortem brain isolated from schizophrenic subjects compared with that of controls (Freedman *et al.*, Biol Psychiatry, 38:22-33, 1995; Court *et al.*, J Neurochem, 73:1590-1597, 1999; and Guan *et al.*, J Neuroreport, 10:1779-1782, 1999).

However, the present invention is the first description of *CHRNA7* alleles associated with decreased  $\alpha 7$  expression. Specifically, the promoter variants in *CHRNA7* identified herein are expected to contribute to the decreased expression of this gene *in vivo*. As described in Example 15 below, several of the polymorphisms have been tested in an *in vitro* reporter gene assay, where 6 of 8 variants were found to have decreased transcriptional activity. In fact, the most common variant at -86 bp, associated with schizophrenia ( $P=0.04$ ), decreased transcription of the luciferase reporter gene by 20% ( $P=0.0001$ ). Comparable transcriptional effects have been seen for other gene promoters (*e.g.*, presenilin 1, tumor necrosis factor, and paraoxonase) with single-base pair mutations (Theuns *et al.*, Hum Mol Genet, 9:325-331, 2000; Knight *et al.*, Nat Genet, 22:145-150, 1999; and Brophy *et al.*, Am J Hum Genet, 68:1428-1436, 2001). Many of the single and double promoter variants identified herein, were found principally in schizophrenic patients. Indeed, the functional variants isolated thus far are statistically more prevalent in schizophrenic subjects ( $P=0.007$ ) than in controls. Additionally, the double variants examined thus far, where more than one variant was present, were combinations of the known single variants, and were found on separate alleles. This indicates inheritance of one mutation from each parent.

It is possible that some variants in the core promoter region have been missed because of ascertainment bias. The sample studied herein included more schizophrenic subjects than controls, and had fewer African Americans and Hispanic subjects than whites. Polymorphisms at -92 bp, -143 bp, -180 bp, and -241 bp were found more often in schizophrenic patients, but were rare in the study sample. Thus, when additional African Americans, Hispanics, and other

ethnic cohorts are screened, more subjects with these rare variants (and possibly even new variants) are expected to be identified. Furthermore, during development of the present invention, an additional 2302 bp of sequence upstream of the *CHRNA7* core promoter was isolated. Preliminary analysis of two subclones indicated the presence of upstream repressor elements. Upstream regulatory elements have been found in several other nicotinic receptor subunit genes indicative of complex developmental and tissue-specific regulation of expression (Flora *et al.*, Eur J Pharmacol, 393:85-95, 2000; and Melnikova *et al.*, Eur J Pharmacol, 393:75-83, 2000). Other functional or more complex variants in schizophrenic subjects are contemplated to lie in these regulatory regions of the human  $\alpha 7$  nicotinic receptor subunit gene, perhaps in disequilibrium with some of the polymorphisms in the core promoter. Although the frequency of core promoter polymorphisms in multiply affected families was small, the polymorphisms associated with decreased  $\alpha 7$  expression are expected to contribute to the transmission of sensory processing deficits in schizophrenia.

Third, because the *CHRNA7* gene was targeted as a candidate gene as having a biological role in a sensory processing endophenotype seen in most schizophrenic patients and in one half of their first degree relatives (Leonard *et al.*, "The role of nicotine and nicotinic receptors in psychopathology," in Arneric and Brioni (eds.), Neuronal Nicotinic Receptors: Pharmacology and Therapeutic Opportunities, New York, NY:Wiley-Liss Inc., pp.305-320, 1998; and Leonard *et al.*, Restro Neurol Neurosci, 12:195-201, 1988), it is significant that a measure of auditory evoked inhibition in humans (the P50 gating phenotype), is correlated with the presence or absence of variants in the *CHRNA7* core promoter. Inhibition of the P50 response is abnormal in most schizophrenic patients, where the test response is often larger than the conditioning response, resulting in T/C ratios much greater than 0.50. In control subjects with no history of schizophrenia, a T/C ratio range lower than in schizophrenic patients was found ( $t_{205}=8.49$ ,  $P<0.0001$ ). However, the ratios were significantly higher in controls with promoter variants than in controls with no polymorphisms ( $P=0.0001$ ). The relationship between the presence of a promoter polymorphism and the P50 T/C ratio appeared to place the control subjects into three groups. The grouping is contemplated to indicate either a gene dosage effect or the presence of additional gene interactions. Inhibitory pathways in schizophrenic subjects are contemplated to be much more complex than in individuals with no history of mental illness. Measurement of the P50 phenotype in control subjects is, thus, contemplated to be less complicated and more

representative of the effect of a few genes or possibly even a single gene defect. The present results indicate that the  $\alpha 7$  promoter variants are associated with a measurable phenotype found in the general population, but present more frequently in schizophrenia. Coincidentally, other investigators have noted correlations between higher P50 ratios and schizotypy (Croft *et al.*, Biol Psychiatry, 50:441-446, 2001), particularly in individuals with a family history of schizophrenia (Cadenhead *et al.*, Am J Psychiatry, 157:1660-1668, 2000). However, none of these investigators had identified or suggested a correlation between *CHRNA7* promoter variants, elevated P50 ratios, and predisposition to schizophrenia.

Last, the design and interpretation of candidate gene association studies, such as the present report, are not obvious. In the human lipoprotein lipase gene, for example, it has been found that the average individual is heterozygous at 17 sites, probably because of a combination of historical population founding, stratification of polymorphic changes, and recombination (Clark *et al.*, Am J Hum Genet, 63:595-612, 1998). Not all of these polymorphisms will be functional, although they may be in disequilibrium with other variants and/or with the disease. This emphasizes the importance of a thorough functional analysis of any polymorphisms associated with schizophrenia. Furthermore, the complexity and dependence on the interactions of functional variants contributing to a complex major mental illness is consistent with the hypothesis that many of these functional polymorphisms are likely to be common in the general population (Lander and Schork, Science, 265:2037-2048, 1990; and Gershon, Biol Psychiatry, 47:240-244, 2000). In that regard, a variant in the catechol O-methyltransferase gene (*COMT*), found in 50% of non-mentally ill subjects, has recently been associated with prefrontal cortical deficits in schizophrenia, but estimated to contribute only a small percentage of the risk for the disease (Egan *et al.*, Proc Natl Acad Sci USA, 98:6917-6922, 2001). Likewise, functional variants in the *CHRNA7* gene promoter were found in 28% of the control subjects with no family history of schizophrenia, but were strongly associated ( $P=0.0001$ ) with having a deficit in auditory sensory processing. The genotype relative risk for schizophrenia at one of the polymorphisms, -86 bp, was 2.39 (95% confidence interval, 1.07-5.32), indicating a small but real contribution to the disorder. This sort of inheritance of gene variants is contemplated to be the case for many complex disorders. Indeed, a role for calpain 10 in type-2 diabetes has been recently reported, where the aberrant allele was found in 75% of the control population but in 80% of those with diabetes (Horikawa *et al.*, Nat Genet, 26:163-175, 2000). Thus, the

assemblage of a group of functional variants in one individual is contemplated to be required for the development of a complex disease such as schizophrenia.

#### IV. Polymorphisms in *CHRNA7* and *dupCHRNA7*

Evidence for genetic linkage to schizophrenia in the 15q13-q14 region has grown as marker density on the human genomic map has improved as described herein and as subsequently published (Stober *et al.*, Am J Hum Gen, 67:1201-1207, 2000; Riley *et al.*, Am J Med Gen, 96:196-201, 2000; Liu *et al.*, Am J Med Gen, 105:658-661, 2001; Tsuang *et al.*, Am J Med Gen, 105:662-668, 2001; and Xu *et al.*, Am J Med Gen, 105:6698-674, 2001). The region has also been linked to bipolar disorder (Turecki *et al.*, Mol. Psych, 6:570-578, 2001). A candidate gene in this region, the  $\alpha 7$  nicotinic acetylcholine receptor subunit gene *CHRNA7*, has been identified pharmacologically, as playing a role in an aberrant inhibitory pathway found in schizophrenia, the P50 auditory evoked potential deficit as described herein and elsewhere (Luntz-Lebman *et al.*, Brain Res, 587:130-136, 1992; Stevens *et al.*, Psychopharm, 136:320-327, 1998; Leonard *et al.*, Pharmacol Biochem Behav, 70:561-570, 2001; and Leonard *et al.*, Eur J Pharmacol, 393:237-242, 2000). The P50 deficit, an endophenotype of schizophrenia, is genetically linked to D15S1360, a dinucleotide marker in intron 2 of *CHRNA7* (Freedman *et al.*, Proc Natl Acad Sci USA, 94:587-592, 1997). Functional gene variants have been isolated in the proximal promoter region of *CHRNA7* that appear to be associated with both schizophrenia and with the P50 deficit as described in Section III herein. Now polymorphisms in the coding region and intron/exon borders of the *CHRNA7* gene cluster in schizophrenic and control subjects are presented.

The mutation screening was complex, due to the partial duplication of the  $\alpha 7$  gene. Exons 5-10, and intervening introns, were duplicated and inserted with a large cassette of DNA into a position proximal to the full-length *CHRNA7*. The duplicated exons are expressed as mRNA with five non- $\alpha 7$  exons in several tissue types, including postmortem brain (*dupCHRNA7*, See, GenBank Accession No. AF029838). Thus, mapping was required, for polymorphisms found in exons 5-10 in genomic DNA, to either the full-length *CHRNA7* gene or the *dupCHRNA7* gene. Transcripts were isolated, specific for each gene, from either postmortem brain tissue or lymphoblasts.

1 Variants in both coding region and introns were identified. As shown in Table 17, 21  
polymorphisms were found the exons, nine of which changed an amino acid. Three of these  
amino acid changes, although rare, mapped to the full-length gene. These three amino acids are  
conserved between human, mouse (GenBank accession #A57175) and rat (GenBank accession  
5 #T01378) genes. One amino acid change in exon 4 (I112V) lies in part of the putative agonist-  
binding site (Galzi *et al.*, Annu Rev Pharmacol, 31:37-72, 1991). In the three families, in which  
these amino acid changes occurred, cosegregation with neither the P50 deficit nor with  
schizophrenia was observed. In such a complex disorder, bilineal inheritance or reduced  
penetrance is contemplated to explain this result. However, functional promoter variants were  
10 found in all three of these families as described herein and as published (Leonard *et al.*, Arch  
Gen Psychiatry, 59:1085-1096, 2002).

Ten intronic variants and two variants in the 3'-untranslated region were identified. Two  
polymorphisms in introns 2 and 3 were in the full-length gene, but the seven variants in introns 7  
and 9 and those in the 3'-untranslated region (3'UT) could not be easily mapped because of the  
15 gene duplication. One variant in intron 9 at +37 was associated with schizophrenia in African  
Americans ( $X^2 = 9.986$ , 1;  $P = 0.0016$ ) and was in linkage disequilibrium with a synonymous  
variant mapped to *CHRNA7*. A number of the intronic polymorphisms either introduce a cryptic  
splice site or alter a splice site. The 2 bp deletion at 497/8, present in the duplicated gene in  
more than 50% of subjects examined, disrupts an exonic splice enhancer site (EXE). Thus, if  
20 exon 6 were aberrantly spliced out in this gene variant, the translation of a putative protein would  
remain in frame, indicating that this splice variant has regulatory effects. Interestingly, multiple  
alternatively spliced transcripts were identified in initial studies of the  $\alpha 7$  gene cluster as  
described in herein. Splice variants have been found to be a common causal element in disease  
(Ars *et al.*, Hum Mol Gen, 9:237-247, 2000; Grabowski and Black, Prog Neurobiol, 65:289-308,  
25 2001; and Cartegni *et al.*, Nat Rev Gen, 3:285-298, 2002). Since the CHRNA7 receptor  
assembles as a pentamer, the presence of splice variants represents a possible mechanism for  
dominant-negative decreased expression (Garcia-Guzman *et al.*, Eur J Neurosci, 7:647-655,  
1995).

The partial duplication of exons 5-10 and flanking regions not only introduced  
30 complexity into the mutation screen, but suggests yet another mechanism of mutation. The  
duplicon containing  $\alpha 7$  exons 5-10 was inserted 3' of five exons, duplicated from another gene,

and the chimera is transcribed in both lymphocytes and brain. This fusion gene or gene product is contemplated to interfere with expression, assembly or function of the *CHRNA7* gene product in a manner similar to a splice variant. Variants in transcribed regions, common to both the *CHRNA7* and *dupCHRNA7* genes were mapped in mRNA from only a limited number of individuals. Thus, it is also contemplated that gene conversion plays a role in disruption of full-length *CHRNA7* in some individuals.

Further, presence of the partial duplication is contemplated to lead to deletion or additional duplication events. For instance, the duplicated sequence is contemplated to prime misalignment, then recombination and subsequent deletion of the intervening sequences including part of the full-length gene. Deletions primed by duplications have been extensively characterized in Prader Willi and Angelman syndromes, which map nearby at 15q11-q13 (Robinson *et al.*, J Med Gen, 35:130-136, 1998). In this regard, five schizophrenic subjects with homozygotic deletions of the duplicated gene have been identified, although none of these subjects appears to be missing any part of the full-length gene. Deletion of both copies of *dupCHRNA7* has not yet been observed in controls.

Although a large number of polymorphisms were found in both the full-length *CHRNA7* gene and its partial duplication, no nucleotide changes that either cosegregate with the P50 gating deficit or schizophrenia, or that obviously disrupt the function of the full-length *CHRNA7* gene were isolated. In addition, none of the coding region variants were found to be in linkage disequilibrium with a functional promoter mutation. Previously, a decreased expression of *CHRNA7* receptors in several regions of postmortem brain in individuals with schizophrenia compared to control subjects has been observed. Since no prominent coding region mutations were found, the promoter polymorphisms described herein in Section III are contemplated to be particularly important, as are intronic variants in the gene. These results also indicate that  $\alpha 7$  nicotinic receptors in schizophrenic subjects, though reduced in number, are functionally normal and thus are contemplated to respond to therapies that modulate  $\alpha 7$  activity or response.

## **V. Detection of *CHRNA7* and *dupCHRNA7* Alleles**

### **A. *CHRNA7* and *dupCHRNA7* Alleles**

In some embodiments, the present invention includes alleles of *CHRNA7* and  
5 *dupCHRNA7* that increase a subject's susceptibility to schizophrenia (*e.g.*, including but not  
limited to alleles containing the promoter variants described herein such as -86C/T, -92G/A, -  
143G/A, -178-G, -180G/C, -191G/A, -194G/C, and -241A/G). Analysis of naturally occurring  
human *CHRNA7* and *dupCHRNA7* alleles revealed that patients with increased susceptibility to  
schizophrenia have a mutant  $\alpha 7$  allele that results in reduced gene transcription and decreased  
10 inhibition in sensory gating (higher P50 T/C ratio). However, the present invention is not limited  
to *CHRNA7* and *dupCHRNA7* alleles with promoter polymorphisms. In fact, any  $\alpha 7$   
polymorphism associated with schizophrenia is within the scope of the present invention. For  
example, in some embodiments, the present invention provides single-nucleotide polymorphisms  
in other regions of *CHRNA7* and *dupCHRNA7* (including but not limited to those shown herein  
15 in Tables 17-19).

### **B. Detection of *CHRNA7* and *dupCHRNA7* Alleles**

Accordingly, the present invention provides methods for determining whether a patient  
has an increased susceptibility to schizophrenia by determining whether the individual has an  $\alpha 7$   
20 allele containing a polymorphism contributing to reduced  $\alpha 7$  expression. In other embodiments,  
the present invention provides methods for providing a prognosis of increased risk for  
schizophrenia to an individual based on the presence or absence of one or more mutations in the  
*CHRNA7* and *dupCHRNA7* genes. In preferred embodiments, the mutation contributes to  
schizophrenia.

25 A number of methods are available for analysis of polymorphisms. Assays for detection  
of polymorphisms or mutations fall into several categories, including, but not limited to direct  
sequencing assays, fragment polymorphism assays, hybridization assays, and computer based  
data analysis. Protocols and commercially available kits or services for performing multiple  
variations of these assays are available. In some embodiments, assays are performed in  
30 combination or in hybrid (*e.g.*, different reagents or technologies from several assays are  
combined to yield one assay). The following assays are useful in the present invention.

## 1. Direct sequencing Assays

In some embodiments of the present invention, polymorphisms are detected using a direct sequencing technique. In these assays, DNA samples are first isolated from a subject using any suitable method. In some embodiments, the region of interest is cloned into a suitable vector and amplified by growth in a host cell (*e.g.*, a bacteria). In other embodiments, DNA in the region of interest is amplified using PCR.

Following amplification, DNA in the region of interest (*e.g.*, the region containing the polymorphism of interest) is sequenced using any suitable method, including but not limited to manual sequencing using radioactive marker nucleotides, or automated sequencing. The results of the sequencing are displayed using any suitable method. The sequence is examined and the presence or absence of a given polymorphism is determined.

## 2. PCR Assay

In some embodiments of the present invention, polymorphisms are detected using a PCR-based assay. In some embodiments, the PCR assay comprises the use of oligonucleotide primers to amplify *CHRNA7* and/or *dupCHRNA7* fragment(s) containing the polymorphism of interest. The presence of an  $\alpha 7$  allele containing nucleotide additions or deletions results in the generation of a longer or shorter PCR fragments respectively, which can be detected by gel electrophoresis.

In other embodiments, the PCR assay comprises the use of oligonucleotide primers that hybridize only to the mutant or wild type allele of  $\alpha 7$  (*e.g.*, to the region of polymorphism). Both sets of primers are used to amplify a sample of DNA. If only the mutant primers result in a PCR product, then the patient has the mutant  $\alpha 7$  allele. If only the wild-type primers result in a PCR product, then the patient has the wild type allele of  $\alpha 7$ .

## 3. Fragment Length Polymorphism Assays

In some embodiments of the present invention, polymorphisms are detected using a fragment length polymorphism assay. In a fragment length polymorphism assay, a unique DNA banding pattern based on cleaving the DNA at a series of positions is generated using an enzyme (*e.g.*, a restriction endonuclease). DNA fragments from a sample containing a polymorphism will have a different banding pattern than wild type.

**a. RFLP Assay**

In some embodiments of the present invention, polymorphisms are detected using a restriction fragment length polymorphism assay (RFLP). The region of interest is first isolated using PCR. The PCR products are then cleaved with restriction enzymes known to give a unique length fragment for a given polymorphism. The restriction-enzyme digested PCR products are separated by agarose gel electrophoresis and visualized by ethidium bromide staining. The length of the fragments is compared to molecular weight markers and fragments generated from wild-type and mutant controls.

**b. CFLP Assay**

In other embodiments, polymorphisms are detected using a CLEAVASE fragment length polymorphism assay (CFLP; Third Wave Technologies, Madison, WI; *See e.g.*, U.S. Patent No.5,888,780). This assay is based on the observation that when single strands of DNA fold on themselves, they assume higher order structures that are highly individual to the precise sequence of the DNA molecule. These secondary structures involve partially duplexed regions of DNA such that single stranded regions are juxtaposed with double stranded DNA hairpins. The CLEAVASE I enzyme, is a structure-specific, thermostable nuclease that recognizes and cleaves the junctions between these single-stranded and double-stranded regions.

The region of interest is first isolated, for example, using PCR. Then, the DNA strands are separated by heating. Next, the reactions are cooled to allow intrastrand secondary structure to form. The PCR products are then treated with the CLEAVASE I enzyme to generate a series of fragments that are unique to a given SNP or mutation. The CLEAVASE enzyme treated PCR products are separated and detected (*e.g.*, by agarose gel electrophoresis) and visualized (*e.g.*, by ethidium bromide staining). The length of the fragments is compared to molecular weight markers and fragments generated from wild-type and mutant controls.

**4. Hybridization Assays**

In preferred embodiments of the present invention, polymorphisms are detected by hybridization assay. In a hybridization assay, the presence or absence of a given polymorphism or mutation is determined based on the ability of the DNA from the sample to hybridize to a

complementary DNA molecule (*e.g.*, a oligonucleotide probe). A variety of hybridization assays using a variety of technologies for hybridization and detection are available. A description of a selection of assays is provided below.

5           **a.       Direct Detection of Hybridization**

In some embodiments, hybridization of a probe to the sequence of interest (*e.g.*, polymorphism) is detected directly by visualizing a bound probe (*e.g.*, a Northern or Southern assay; *See e.g.*, Ausabel *et al.* (eds.), Current Protocols in Molecular Biology, John Wiley & Sons, NY, 1991). In these assays, genomic DNA (Southern) or RNA (Northern) is isolated from  
10 a subject. The DNA or RNA is then cleaved with a series of restriction enzymes that cleave infrequently in the genome and not near any of the markers being assayed. The DNA or RNA is then separated (*e.g.*, agarose gel electrophoresis) and transferred to a membrane. A labeled (*e.g.*, by incorporating a radionucleotide) probe or probes specific for the mutation being detected is allowed to contact the membrane under a condition of low, medium, or high stringency  
15 conditions. Unbound probe is removed and the presence of binding is detected by visualizing the labeled probe.

**b.       Detection of Hybridization Using "DNA Chip" Assays**

In some embodiments of the present invention, polymorphisms are detected using a DNA  
20 chip hybridization assay. In this assay, a series of oligonucleotide probes are affixed to a solid support. The oligonucleotide probes are designed to be unique to a given polymorphism. The DNA sample of interest is contacted with the DNA "chip" and hybridization is detected.

In some embodiments, the DNA chip assay is a GeneChip (Affymetrix, Santa Clara, CA; *See e.g.*, U.S. Patent No. 6,045,996) assay. The GeneChip technology uses miniaturized,  
25 high-density arrays of oligonucleotide probes affixed to a "chip." Probe arrays are manufactured by Affymetrix's light-directed chemical synthesis process, which combines solid-phase chemical synthesis with photolithographic fabrication techniques employed in the semiconductor industry. Using a series of photolithographic masks to define chip exposure sites, followed by specific chemical synthesis steps, the process constructs high-density arrays of oligonucleotides, with  
30 each probe in a predefined position in the array. Multiple probe arrays are synthesized simultaneously on a large glass wafer. The wafers are then diced, and individual probe arrays

are packaged in injection-molded plastic cartridges, which protect them from the environment and serve as chambers for hybridization.

The nucleic acid to be analyzed is isolated, amplified by PCR, and labeled with a fluorescent reporter group. The labeled DNA is then incubated with the array using a fluidics station. The array is then inserted into the scanner, where patterns of hybridization are detected. The hybridization data are collected as light emitted from the fluorescent reporter groups already incorporated into the target, which is bound to the probe array. Probes that perfectly match the target generally produce stronger signals than those that have mismatches. Since the sequence and position of each probe on the array are known, by complementarity, the identity of the target nucleic acid applied to the probe array can be determined.

In other embodiments, a DNA microchip containing electronically captured probes (Nanogen, San Diego, CA) is utilized (*See e.g.*, U.S. Patent No. 6,068,818). Through the use of microelectronics, Nanogen's technology enables the active movement and concentration of charged molecules to and from designated test sites on its semiconductor microchip. DNA capture probes unique to a given SNP or mutation are electronically placed at, or "addressed" to, specific sites on the microchip. Since DNA has a strong negative charge, it can be electronically moved to an area of positive charge.

First, a test site or a row of test sites on the microchip is electronically activated with a positive charge. Next, a solution containing the DNA probes is introduced onto the microchip. The negatively charged probes rapidly move to the positively charged sites, where they concentrate and are chemically bound to a site on the microchip. The microchip is then washed and another solution of distinct DNA probes is added until the array of specifically bound DNA probes is complete.

A test sample is then analyzed for the presence of target DNA molecules by determining which of the DNA capture probes hybridize, with complementary DNA in the test sample (*e.g.*, a PCR amplified gene of interest). An electronic charge is also used to move and concentrate target molecules to one or more test sites on the microchip. The electronic concentration of sample DNA at each test site promotes rapid hybridization of sample DNA with complementary capture probes (hybridization may occur in minutes). To remove any unbound or nonspecifically bound DNA from each site, the polarity or charge of the site is reversed to negative, thereby

forcing any unbound or nonspecifically bound DNA back into solution away from the capture probes. A laser-based fluorescence scanner is used to detect binding,

In still further embodiments, an array technology based upon the segregation of fluids on a flat surface (chip) by differences in surface tension (ProtoGene, Palo Alto, CA) is utilized (*See e.g.*, U.S. Patent No. 6,001,311). Protogene's technology is based on the fact that fluids can be segregated on a flat surface by differences in surface tension that have been imparted by chemical coatings. Once so segregated, oligonucleotide probes are synthesized directly on the chip by ink-jet printing of reagents. The array with its reaction sites defined by surface tension is mounted on a X/Y translation stage under a set of four piezoelectric nozzles, one for each of the four standard DNA bases. The translation stage moves along each of the rows of the array and the appropriate reagent is delivered to each of the reaction site. For example, the A amidite is delivered only to the sites where amidite A is to be coupled during that synthesis step and so on. Common reagents and washes are delivered by flooding the entire surface, and are then removed by spinning.

DNA probes unique for the polymorphism of interest are affixed to the chip using Protogene's technology. The chip is then contacted with the PCR-amplified genes of interest. Following hybridization, unbound DNA is removed and hybridization is detected using any suitable method (*e.g.*, by fluorescence de-quenching of an incorporated fluorescent group).

In yet other embodiments, a "bead array" is used for the detection of polymorphisms (Illumina, San Diego, CA; *See e.g.*, PCT Publications WO 99/67641 and WO 00/39587, each of which is herein incorporated by reference). Illumina uses a BEAD ARRAY technology that combines fiber optic bundles and beads that self-assemble into an array. Each fiber optic bundle contains thousands to millions of individual fibers depending on the diameter of the bundle. The beads are coated with an oligonucleotide specific for the detection of a given SNP or mutation. Batches of beads are combined to form a pool specific to the array. To perform an assay, the BEAD ARRAY is contacted with a prepared subject sample (*e.g.*, DNA). Hybridization is detected using any suitable method.

### **c. Enzymatic Detection of Hybridization**

In some embodiments of the present invention, genomic profiles are generated using a assay that detects hybridization by enzymatic cleavage of specific structures (INVADER assay,

Third Wave Technologies; *See e.g.*, U.S. Patent No. 6,001,567). The INVADER assay detects specific DNA and RNA sequences by using structure-specific enzymes to cleave a complex formed by the hybridization of overlapping oligonucleotide probes. Elevated temperature and an excess of one of the probes enable multiple probes to be cleaved for each target sequence present without temperature cycling. These cleaved probes then direct cleavage of a second labeled probe. The secondary probe oligonucleotide can be 5'-end labeled with fluorescein that is quenched by an internal dye. Upon cleavage, the de-quenched fluorescein labeled product may be detected using a standard fluorescence plate reader.

The INVADER assay detects specific mutations and SNPs in unamplified genomic DNA. The isolated DNA sample is contacted with the first probe specific either for a SNP/mutation or wild type sequence and allowed to hybridize. Then a secondary probe, specific to the first probe, and containing the fluorescein label, is hybridized and the enzyme is added. Binding is detected by using a fluorescent plate reader to compare the signal of the test sample to known positive and negative controls.

In some embodiments, hybridization of a bound probe is detected using a TaqMan assay (PE Biosystems, Foster City, CA; *See e.g.*, U.S. Patent No. 5,962,233). The assay is performed during a PCR reaction. The TaqMan assay exploits the 5'-3' exonuclease activity of the AMPLITAQ GOLD DNA polymerase. A probe, specific for a given allele or mutation, is included in the PCR reaction. The probe consists of an oligonucleotide with a 5'-reporter dye (*e.g.*, a fluorescent dye) and a 3'-quencher dye. During PCR, if the probe is bound to its target, the 5'-3' nucleolytic activity of the AMPLITAQ GOLD polymerase cleaves the probe between the reporter and the quencher dye. The separation of the reporter dye from the quencher dye results in an increase of fluorescence. The signal accumulates with each cycle of PCR and can be monitored with a fluorimeter.

In still further embodiments, polymorphisms are detected using the SNP-IT primer extension assay (Orchid Biosciences, Princeton, NJ; *See e.g.*, U.S. Patent No. 5,952,174). SNPs are identified in this assay, by using a specially synthesized DNA primer and a DNA polymerase, to selectively extend the DNA chain by one base at the suspected SNP location. DNA in the region of interest is amplified and denatured. Polymerase reactions are then performed using miniaturized systems called microfluidics. Detection is accomplished by adding a label to the nucleotide suspected of being at the SNP or mutation location.

Incorporation of the label into the DNA can be detected by any suitable method (*e.g.*, if the nucleotide contains a biotin label, detection is via a fluorescently labeled antibody specific for biotin).

## 5. Mass Spectroscopy Assay

In some embodiments, a MassARRAY system (Sequenom, San Diego, CA.) is used to detect polymorphisms (*See e.g.*, U.S. Patent No. 6,043,031). DNA is isolated from blood samples using standard procedures. Next, specific DNA regions containing the polymorphism of interest, about 200 base pairs in length, are amplified by PCR. The amplified fragments are then attached by one strand to a solid surface and the non-immobilized strands are removed by standard denaturation and washing. The remaining immobilized single strand then serves as a template for automated enzymatic reactions that produce genotype specific diagnostic products.

Very small quantities of the enzymatic products, typically five to ten nanoliters, are then transferred to a SpectroCHIP array for subsequent automated analysis with the SpectroREADER mass spectrometer. Each spot is preloaded with light absorbing crystals that form a matrix with the dispensed diagnostic product. The MassARRAY system uses MALDI-TOF (Matrix Assisted Laser Desorption Ionization - Time of Flight) mass spectrometry. In a process known as desorption, the matrix is hit with a pulse from a laser beam. Energy from the laser beam is transferred to the matrix and it is vaporized resulting in a small amount of the diagnostic product being expelled into a flight tube. As the diagnostic product is charged when an electrical field pulse is subsequently applied to the tube they are launched down the flight tube towards a detector. The time between application of the electrical field pulse and collision of the diagnostic product with the detector is referred to as the time of flight. This is a very precise measure of the product's molecular weight, as a molecule's mass correlates directly with time of flight with smaller molecules flying faster than larger molecules. The entire assay is completed in less than one thousandth of a second, enabling samples to be analyzed in a total of 3-5 second including repetitive data collection. The SpectroTYPER software then calculates, records, compares and reports the genotypes at the rate of three seconds per sample.

## 6. Kits for Analyzing Risk of Schizophrenia

The present invention also provides kits for determining whether an individual possesses an  $\alpha 7$  allele with a specific polymorphism. In some embodiments, the kits are useful in determining whether the subject is at risk of developing schizophrenia. The diagnostic kits are produced in a variety of ways. In some embodiments, the kits contain at least one reagent for specifically detecting a mutant  $\alpha 7$  allele. In preferred embodiments, the kits contain reagents for detecting polymorphisms in the  $\alpha 7$  gene promoter. In preferred embodiments, the reagents are primers for amplifying the region of DNA containing the promoter. In other preferred embodiments, the reagent is a probe that binds to the polymorphic region. In some  
10   embodiments, the kit contains instructions for determining whether the subject is at risk for developing schizophrenia. In preferred embodiments, the instructions specify that risk for developing schizophrenia is determined by detecting the presence or absence of a mutant  $\alpha 7$  allele in the subject, wherein subjects having an allele containing a promoter polymorphism associated with decreased  $\alpha 7$  transcription, have an increased risk of developing schizophrenia.  
15   In some embodiments, the kits include ancillary reagents such as buffering agents, nucleic acid stabilizing reagents, protein stabilizing reagents, and signal producing systems (*e.g.*, fluorescence generating systems). The test kit may be packaged in any suitable manner, typically with the elements in a single container or various containers as necessary along with a sheet of instructions for carrying out the test. In some embodiments, the kits also preferably include a  
20   positive control sample.

## 7. Bioinformatics

In some embodiments, the present invention provides methods of determining an individual's risk of developing schizophrenia based on the presence of one or more mutant alleles  
25   of  $\alpha 7$ . In some embodiments, the analysis of polymorphism data is automated. For example, in some embodiments, the present invention provides a bioinformatics research system comprising a plurality of computers running a mullet-platform object oriented programming language (*See e.g.*, U.S. Patent 6,125,383). In some embodiments, one of the computers stores genetics data (*e.g.*, the risk of contracting schizophrenia associated with a given polymorphism). In some  
30   embodiments, one of the computers stores application programs (*e.g.*, for analyzing transmission

disequilibrium data or determining genotype relative risks and population attributable risks). Results are then delivered to the user (e.g., via one of the computers or via the internet).

## **VI. Treatment and Diagnosis of Schizophrenia and Other Psychoses**

5 The present invention provides methods and compositions for the development and identification of alternative means to diagnose and treat schizophrenia. The methods and compositions of the present invention will find use in the functional assessment of  $\alpha 7$  nicotinic receptors in schizophrenic patients, as well as for screening populations for deficits in receptor function. The present invention finds use in genetic screening methods for genetic and parentage  
10 counseling, as well as for identification of individuals at risk for developing schizophrenia.

The present invention also provides methods and compositions for modifying  $\alpha 7$  nicotinic receptor function. For example, the present invention contemplates the development of genetic therapy methods to correct deficiencies in the receptor structure and/or function, as well as other therapeutic methods to enhance or decrease the function of the receptor, as appropriate  
15 for the treatment of any given individual.

It is also contemplated that the present invention will find use in relation to other psychosis. For example, the present invention will find use in the diagnosis and treatment of genetic disorders, in particular those genetic disorders known to have a genetic component associated with chromosome 15, such as Prader-Willi syndrome, Angelman's syndrome, etc., as  
20 well as other diseases, such as epilepsy (e.g., juvenile myoclonic epilepsy), breast, and other types of cancers. The present invention also finds use in the diagnosis and treatment of nicotine-dependent illnesses, including, but not limited, to small cell lung carcinoma.

Indeed, it is contemplated that the present invention will find use in the development of antipsychotic drugs targeted to the  $\alpha 7$  nicotinic receptor and/or the  $\alpha 7$  nicotinic receptor subunit  
25 gene. For example, dimethylbenzylidene anabaseine (DMXB-A; [(2-4) Dimethoxy-benzylidene anabaseine hydrochloride]) and its congeners are selectively agonistic at the  $\alpha 7$  receptor. During the development of the present invention, an animal model of the deficit observed in schizophrenics was used to show that DMXB-A is effective in repeated doses, whereas the effect of nicotine itself is completely inactivated after one dose. DMXB-A also has significantly less  
30 cardiovascular effects than nicotine, consistent with its antagonist effects at  $\alpha 4$ - $\alpha 2$  nicotinic receptors. Thus, it is contemplated that DMXB-A will find use as an anti-psychotic drug.

In addition to the physiological deficit found in schizophrenics and some of their relatives, similar deficits are also found as state-related changes in other psychotic disorders, including Parkinson's, Alzheimer's, mania and cocaine dependence. In stimulant dependence, neuroleptic anti-psychotic drugs have poor patient compliance, possibly due to their anhedonic, catecholamine-blocking effects. Thus, it is contemplated that nicotinic cholinergic therapeutic strategies, such as those developed using the methods and/or compositions of the present invention will be effective against a broad spectrum of clinical indications.

It is further contemplated that the present invention will be used to develop antibodies and other diagnostic reagents. For example, the present invention finds use in the production of peptide antibodies using sequences identified using the present invention.

## **VII. Transgenic Animals**

The present invention provides methods and compositions for production of transgenic animal models of schizophrenia, nicotine-dependent illnesses, and cancer. It is also contemplated that such systems as *Xenopus* oocytes will be used to express human  $\alpha 7$  receptors and gene sequences of the present invention.

In preferred embodiments, transgenic mice are generated using microinjection of DNA containing  $\alpha 7$  gene sequences into mammalian oocytes. However, equivalent transgenic mice can also be produced by homologous recombination in embryonic stem (ES) cells. Techniques for the isolation, culture, microinjection and implantation of a variety of mammalian oocytes (e.g., murine, porcine, ovine, bovine, etc.) are known to the art.

Two mouse models are provided in the present invention. The first model involves introduction of an intact human  $\alpha 7$  gene into the mouse genome by microinjection of a fertilized egg with DNA from the clone containing the full-length nAChR gene described in Example 8. The integrity of the clone in the transgenic mice is examined by PCR amplification, using all of the identified STSs on the clone map. Large flanking DNA sequences are included in this transgene, in order to ensure proper expression of the human  $\alpha 7$  gene in the mice. The expression of the human  $\alpha 7$  gene in mice is examined by an RNase protection assay designed to specifically detect the human  $\alpha 7$  mRNA. This expression pattern coincides with the expression pattern of  $\alpha 7$  in human tissues, as analyzed by Northern hybridization. The transgenic mouse model provides animals for determinations of  $\alpha 7$  function in nicotine-dependence, nicotine-

dependent illnesses, cancers associated with chromosome 15, schizophrenia, and other psychoses. These animals also facilitate the development of drugs and other therapeutics that affect the function of human  $\alpha 7$  *in vivo*.

The second model is exemplified using transgenic mice that contain targeted disruptions of the  $\alpha 7$  gene. These animals, termed "knockout" animals, lack the ability to express  $\alpha 7$  (" $\alpha 7$  knockouts"). In this model, mice are generated with a deletion specifically in the  $\alpha 7$  gene, in order to allow assessment of phenotypic changes. In order to produce the transgenic knockout mice of the present invention, cloned human  $\alpha 7$  gene sequences are used to disrupt the  $\alpha 7$  gene in such a manner that  $\alpha 7$  cannot be produced. In this model, two types of deletions are designed. The first removes the  $\alpha 7$  gene entirely from the germline cells. The second type of deletion is engineered so as to provide control over the specific tissue and developmental stage in which  $\alpha 7$  expression is interrupted. In the second model, the viability of the mutated animals is maintained, permitting analysis of the animals' phenotypes (including expression in specific tissues).

## DEFINITIONS

To facilitate understanding of the invention, a number of terms are defined below.

The term " $\alpha 7$  gene" (or *Alpha-7*, or "Alpha-7 gene") refers to the full-length  $\alpha 7$  nucleotide sequence. However, it is also intended that the term encompass fragments of the  $\alpha 7$  sequence, such as those set forth as SEQ ID NOS:94, 101, 122, and 125, as well as other domains within the full-length  $\alpha 7$  nucleotide sequence. Furthermore, the terms "Alpha-7 nucleotide sequence" or "Alpha-7 polynucleotide sequence" (or " $\alpha 7$  nucleotide sequence" or " $\alpha 7$  polynucleotide sequence") encompasses DNA, cDNA, and RNA (*e.g.*, mRNA) sequences. In preferred embodiments, the  $\alpha 7$  is human  $\alpha 7$ .

A "variant" of human  $\alpha 7$  as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, (*e.g.*, replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (*e.g.*, replacement of a glycine with a tryptophan). Similar minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted,

inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

The term “biologically active,” as used herein, refers to a protein or other biologically active molecules (*e.g.*, catalytic RNA) having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, “immunologically active” refers to the capability of the natural, recombinant, or synthetic  $\alpha 7$ , or any oligopeptide or polynucleotide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term “agonist,” as used herein, refers to a molecule which, when bound to  $\alpha 7$ , causes a change in  $\alpha 7$ , which modulates the activity of  $\alpha 7$ . Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules that bind to or interact with  $\alpha 7$ .

The terms “antagonist” or “inhibitor,” as used herein, refer to a molecule which, when bound to  $\alpha 7$ , blocks or modulates the biological or immunological activity of  $\alpha 7$ . Antagonists and inhibitors may include proteins, nucleic acids, carbohydrates, or any other molecules, which bind or interact with  $\alpha 7$ .

The term “modulate,” as used herein, refers to a change or an alteration in the biological activity of  $\alpha 7$ . Modulation may be an increase or a decrease in protein activity, a change in binding characteristics, or any other change in the biological, functional, or immunological properties of  $\alpha 7$ .

The term “derivative,” as used herein, refers to the chemical modification of a nucleic acid encoding  $\alpha 7$ , or the encoded  $\alpha 7$ . Illustrative of such modifications would be replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative would encode a polypeptide that retains essential biological characteristics of the natural molecule.

The term “gene” refers to a nucleic acid (*e.g.*, DNA) sequence that comprises coding sequences necessary for the production of a polypeptide or precursor (*e.g.*, human  $\alpha 7$ ). The polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired activity or functional properties (*e.g.*, enzymatic activity, ligand binding, signal transduction, etc.) of the full-length or fragment are retained. The term also encompasses the coding region of a structural gene and the including sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb on either end such that

the gene corresponds to the length of the full-length mRNA. The sequences which are located 5' of the coding region and which are present on the mRNA are referred to as 5' non-translated sequences. The sequences which are located 3' or downstream of the coding region and which are present on the mRNA are referred to as 3' non-translated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene.

A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene that are transcribed into nuclear RNA (hnRNA). Introns may contain regulatory elements such as enhancers. Introns are removed or "spliced out" from the nuclear or primary transcript, and thus introns are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide.

Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms, such as "polypeptide" or "protein," are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5' and 3' end of the sequences that are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers that control or influence the transcription of the gene. The 3' flanking region may contain sequences that direct the termination of transcription, post-transcriptional cleavage and polyadenylation. The present invention provides DNA sequence of the  $\alpha 7$  promoter (SEQ ID NO:101; *See*, Figure 8). The present invention also provides DNA sequence for the region located 5' to the human  $\alpha 7$  gene (SEQ ID NO:94; *See*, Figure 4).

"Peptide nucleic acid," as used herein, refers to a molecule that comprises an oligomer to which an amino acid residue, such as lysine, and an amino group have been added. These small molecules, also designated anti-gene agents, stop transcript elongation by binding to their complementary strand of nucleic acid (Nielsen *et al.*, *Anticancer Drug Des*, 8:53-63, 1993).

The term "wild-type" refers to a gene or gene product that has the characteristics of that gene or gene product when isolated from a naturally occurring source. A wild-type gene is that which is most frequently observed in a population and is thus arbitrarily designed the "normal" or "wild-type" form of the gene. In contrast, the term "modified" or "mutant" refers to a gene or gene product that displays modifications in sequence and or functional properties (*i.e.*, altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product.

As used herein, the term "polymorphism" refers to the regular and simultaneous occurrence in a single interbreeding population of two or more alleles of a gene, where the frequency of the rarer allele(s) is greater than can be explained by recurrent mutation alone (typically greater than 1%). In preferred embodiments, the term "polymorphism" refers to at least one substitution, insertion and/or deletion in the 5' untranslated region of  $\alpha 7$ . In particularly preferred embodiments, the polymorphism is in the  $\alpha 7$  promoter and contributes to a reduction in  $\alpha 7$  transcription. In other preferred embodiments, the polymorphism is associated with a predisposition to schizophrenia.

The term "allele" refers to one of at least two mutually exclusive forms of the same gene, occupying the same locus on homologous chromosomes, and governing the same biochemical and developmental process.

As used herein, the terms "nucleic acid molecule encoding," "DNA sequence encoding," and "DNA encoding" refer to the order or sequence of deoxyribonucleotides along a strand of deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along the polypeptide (protein) chain. The DNA sequence thus codes for the amino acid sequence.

DNA molecules are said to have "5' ends" and "3' ends" because mononucleotides are reacted to make oligonucleotides or polynucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage. Therefore, an end of an oligonucleotide or a polynucleotide, referred to as the "5' end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a

larger oligonucleotide or polynucleotide, also may be said to have 5' and 3' ends. In either a linear or circular DNA molecule, discrete elements are referred to as being "upstream" or 5' of the "downstream" or 3' elements. This terminology reflects the fact that transcription proceeds in a 5' to 3' fashion along the DNA strand. The promoter and enhancer elements which direct transcription of a linked gene are generally located 5' or upstream of the coding region. However, enhancer elements can exert their effect even when located 3' of the promoter element and the coding region. Transcription termination and polyadenylation signals are located 3' or downstream of the coding region.

As used herein, the terms "an oligonucleotide having a nucleotide sequence encoding a gene" and "polynucleotide having a nucleotide sequence encoding a gene," means a nucleic acid sequence comprising the coding region of a gene or in other words the nucleic acid sequence that encodes a gene product. The coding region may be present in either a cDNA, a genomic DNA or an RNA form. When present in a DNA form, the oligonucleotide or polynucleotide may be single-stranded (*i.e.*, the sense strand) or double-stranded. Suitable control elements such as enhancers/promoters, splice junctions, polyadenylation signals, *etc.* may be placed in close proximity to the coding region of the gene if needed to permit proper initiation of transcription and/or correct processing of the primary RNA transcript. Alternatively, the coding region utilized in the expression vectors of the present invention may contain endogenous enhancers/promoters, splice junctions, intervening sequences, polyadenylation signals, *etc.* or a combination of both endogenous and exogenous control elements.

As used herein, the term "regulatory element" refers to a genetic element that controls some aspect of the expression of nucleic acid sequences. For example, a promoter is a regulatory element that facilitates the initiation of transcription of an operably linked coding region. Other regulatory elements are splicing signals, polyadenylation signals, termination signals, etcetera.

Transcriptional control signals in eukaryotes comprise "promoter" and "enhancer" elements, and in some cases further comprise operator sequences. Promoters and enhancers consist of short arrays of DNA sequences that interact specifically with cellular proteins involved in transcription (Maniatis *et al.*, Science, 236:1237, 1987). Promoter and enhancer elements have been isolated from a variety of eukaryotic sources including genes in yeast, insect and mammalian cells and viruses (analogous control elements, *i.e.*, promoters, are also found in prokaryote). The selection of a particular promoter and enhancer depends on what cell type is to

be used to express the protein of interest. Some eukaryotic promoters and enhancers have a broad host range while others are functional in a limited subset of cell types (Voss *et al.*, Trends Biochem Sci, 11:287, 1986; and Maniatis *et al.*, *supra*, 1987). For example, the SV40 early gene enhancer is very active in a wide variety of cell types from many mammalian species and has  
5 been widely used for the expression of proteins in mammalian cells (Dijkema *et al.*, EMBO J, 4:761, 1985). Two other examples of promoter/enhancer elements active in a broad range of mammalian cell types are those from the human elongation factor 1 $\alpha$  gene (Uetsuki *et al.*, J Biol Chem, 264:5791, 1989; Kim *et al.*, Gene 91:217, 1990; and Mizushima and Nagata, Nuc Acids Res, 18:5322, 1990) and the long terminal repeats of the Rous sarcoma virus (Gorman *et al.*,  
10 Proc Natl Acad Sci USA, 79:6777, 1982) and the human cytomegalovirus (Boshart *et al.*, Cell, 41:521, 1985).

As used herein, the term "promoter/enhancer" denotes a segment of DNA which contains sequences capable of providing both promoter and enhancer functions (*i.e.*, the functions provided by a promoter element and an enhancer element, see above for a discussion of these  
15 functions). For example, the long terminal repeats of retroviruses contain both promoter and enhancer functions. The enhancer/promoter may be "endogenous" or "exogenous" or "heterologous." An "endogenous" enhancer/promoter is one that is naturally linked with a given gene in the genome. An "exogenous" or "heterologous" enhancer/promoter is one that is placed in juxtaposition to a gene by means of genetic manipulation (*i.e.*, molecular biological  
20 techniques) such that transcription of that gene is directed by the linked enhancer/promoter.

The term "repressor" as used herein refers to a regulatory protein that binds to an operator of a gene to prevent transcription of the gene. The binding affinity of repressors for the operator may be affected by other molecules. Inducers bind to repressors and decrease their binding to the operator, while corepressors increase the binding. As used herein, the terms  
25 "operator" and "repressor sequence" refer to the site on DNA to which a specific repressor protein binds thereby preventing the initiation of transcription at the adjacent promoter.

The presence of "splicing signals" on an expression vector often results in higher levels of expression of the recombinant transcript. Splicing signals mediate the removal of introns from the primary RNA transcript and consist of a splice donor and acceptor site (Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press,  
30

New York, pp. 16.6-16.8, 1989). A commonly used splice donor and acceptor site is the splice junction from the 16S RNA of SV40.

Efficient expression of recombinant DNA sequences in eukaryotic cells requires expression of signals directing the efficient termination and polyadenylation of the resulting transcript. Transcription termination signals are generally found downstream of the polyadenylation signal and are a few hundred nucleotides in length. The term "polyA site" or "polyA sequence" as used herein denotes a DNA sequence, which directs both the termination and polyadenylation of the nascent RNA transcript. Efficient polyadenylation of the recombinant transcript is desirable as transcripts lacking a polyA tail are unstable and are rapidly degraded. The polyA signal utilized in an expression vector may be "heterologous" or "endogenous." An endogenous polyA signal is one that is found naturally at the 3' end of the coding region of a given gene in the genome. A heterologous polyA signal is one that is isolated from one gene and placed 3' of another gene. A commonly used heterologous polyA signal is the SV40 polyA signal. The SV40 polyA signal is contained on a 237 bp *Bam*HI/*Bcl*II restriction fragment and directs both termination and polyadenylation (Sambrook, *supra*, 1989).

Eukaryotic expression vectors may also contain "viral replicons" or "viral origins of replication." Viral replicons are viral DNA sequences that allow for the extrachromosomal replication of a vector in a host cell expressing the appropriate replication factors. Vectors that contain either the SV40 or polyoma virus origin of replication replicate to high copy number (up to  $10^4$  copies/cell) in cells that express the appropriate viral T antigen. Vectors that contain the replicons from bovine papillomavirus or Epstein-Barr virus replicate extrachromosomally at low copy number (~100 copies/cell).

As used herein, the terms "complementary" or "complementarity" are used in reference to polynucleotides (*i.e.*, a sequence of nucleotides) related by the base-pairing rules. For example, for the sequence "A-G-T," is complementary to the sequence "T-C-A." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods that depend upon binding between nucleic acids.

The term "homology" refers to a degree of complementarity. There may be partial homology or complete homology (*i.e.*, identity). A partially complementary sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (*i.e.*, the hybridization) of a completely homologous to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (*i.e.*, selective) interaction. The absence of non-specific binding may be tested by the use of a second target which lacks even a partial degree of complementarity (*e.g.*, less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

The art knows well that numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (*e.g.*, the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, the art knows conditions that promote hybridization under conditions of high stringency (*e.g.*, increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.).

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe which can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described above.

A gene may produce multiple RNA species, which are generated by differential splicing of the primary RNA transcript. cDNAs that are splice variants of the same gene will contain regions of sequence identity or complete homology (representing the presence of the same exon or portion of the same exon on both cDNAs) and regions of complete non-identity (for example,

representing the presence of exon "A" on cDNA 1 wherein cDNA 2 contains exon "B" instead). Because the two cDNAs contain regions of sequence identity they will both hybridize to a probe derived from the entire gene or portions of the gene containing sequences found on both cDNAs; the two splice variants are therefore substantially homologous to such a probe and to each other.

5 When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe which can hybridize (*i.e.*, it is the complement of) the single-stranded nucleic acid sequence under conditions of low stringency as described above.

The term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the  $T_m$  of the formed hybrid, and the G:C ratio within the nucleic acids. A single molecule that contains pairing of complementary nucleic acids within its structure is said to be "self-hybridized."

15 As used herein, the term " $T_m$ " is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the  $T_m$  of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the  $T_m$  value may be calculated by the equation:  $T_m = 81.5 + 0.41(\% G + C)$ , when a nucleic acid is in aqueous solution at 1 M NaCl (*See e.g.*, Anderson and Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization, 1985). Other references include more sophisticated computations that take structural as well as sequence characteristics into account for the calculation of  $T_m$ .

25 As used herein the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. Under "low stringency conditions" a nucleic acid sequence of interest will hybridize to its exact complement, sequences with single base mismatches, closely related sequences (*e.g.*, sequences with 90% or greater homology), and sequences having only partial homology (*e.g.*, sequences with 50-90% homology). Under 30 'medium stringency conditions,' a nucleic acid sequence of interest will hybridize only to its

exact complement, sequences with single base mismatches, and closely related sequences (e.g., 90% or greater homology). Under "high stringency conditions," a nucleic acid sequence of interest will hybridize only to its exact complement, and (depending on conditions such as temperature) sequences with single base mismatches. In other words, under conditions of high stringency the temperature can be raised so as to exclude hybridization to sequences with single base mismatches.

"High stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH<sub>2</sub>PO<sub>4</sub> H<sub>2</sub>O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1X SSPE, 1.0% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

"Medium stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH<sub>2</sub>PO<sub>4</sub> H<sub>2</sub>O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 1.0X SSPE, 1.0% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

"Low stringency conditions" comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH<sub>2</sub>PO<sub>4</sub> H<sub>2</sub>O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5X Denhardt's reagent [50X Denhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V; Sigma)] and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 5X SSPE, 0.1% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

The art knows well that numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, the art knows conditions that promote

hybridization under conditions of high stringency (*e.g.*, increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.) (see definition above for "stringency").

"Amplification" is a special case of nucleic acid replication involving template specificity. It is to be contrasted with non-specific template replication (*i.e.*, replication that is template-dependent but not dependent on a specific template). Template specificity is here distinguished from fidelity of replication (*i.e.*, synthesis of the proper polynucleotide sequence) and nucleotide (ribo- or deoxyribo-) specificity. Template specificity is frequently described in terms of "target" specificity. Target sequences are "targets" in the sense that they are sought to be sorted out from other nucleic acid. Amplification techniques have been designed primarily for this sorting out.

Template specificity is achieved in most amplification techniques by the choice of enzyme. Amplification enzymes are enzymes that, under conditions they are used, will process only specific sequences of nucleic acid in a heterogeneous mixture of nucleic acid. For example, in the case of Q $\beta$  replicase, MDV-1 RNA is the specific template for the replicase (Kacian *et al.*, Proc Natl Acad Sci USA, 69:3038, 1972). This amplification enzyme does not replicate other nucleic acids. Similarly, in the case of T7 RNA polymerase, this amplification enzyme has a stringent specificity for its own promoters (Chamberlin *et al.*, Nature, 228:227, 1970). In the case of T4 DNA ligase, the enzyme will not ligate the two oligonucleotides or polynucleotides, where there is a mismatch between the oligonucleotide or polynucleotide substrate and the template at the ligation junction (Wu and Wallace, Genomics, 4:560, 1989). Finally, *Taq* and *Pfu* polymerases, by virtue of their ability to function at high temperature, are found to display high specificity for the sequences bounded and thus defined by the primers; the high temperature results in thermodynamic conditions that favor primer hybridization with the target sequences and not hybridization with non-target sequences (Erich (ed.), *PCR Technology*, Stockton Press, 1989).

As used herein, the term "amplifiable nucleic acid" is used in reference to nucleic acids that may be amplified by any amplification method. It is contemplated that "amplifiable nucleic acid" will usually comprise "sample template."

As used herein, the term "sample template" refers to nucleic acid originating from a sample, which is analyzed for the presence of "target" (defined below). In contrast, "background

template" is used in reference to nucleic acid other than sample template, which may or may not be present in a sample. Background template is most often inadvertent. It may be the result of carryover, or it may be due to the presence of nucleic acid contaminants sought to be purified away from the sample. For example, nucleic acids from organisms other than those to be  
5 detected may be present as background in a test sample.

As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (*i.e.*, in the  
10 presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the  
15 synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method. The present invention provides sequences for numerous primers (*i.e.*, SEQ ID NOS:1-8, and 12-83).

The term "sense primer" refers to an oligonucleotide capable of hybridizing to the  
20 noncoding strand of gene. The term "antisense primer" refers to an oligonucleotide capable of hybridizing to the coding strand of a gene.

As used herein, the term "fluorescent tag" refers to a molecule having the ability to emit light of a certain wavelength when activated by light of another wavelength. "Fluorescent tags" suitable for use with the present invention include but are not limited to fluorescein, rhodamine,  
25 Texas red, 6-FAM, TET, HEX, Cy5, Cy3, and Oregon Green.

As used herein, the term "probe" refers to an oligonucleotide (*i.e.*, a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, which is capable of hybridizing to another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are  
30 useful in the detection, identification and isolation of particular gene sequences. It is contemplated that any probe used in the present invention will be labeled with any "reporter

molecule," so that is detectable in any detection system, including, but not limited to enzyme (e.g., ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label. The present invention provides sequences for suitable for use as probes (e.g., SEQ ID NO:9-11, as well as the primer sequences described above).

As used herein, the term "target," when used in reference to the polymerase chain reaction, refers to the region of nucleic acid bounded by the primers used for polymerase chain reaction. Thus, the "target" is sought to be sorted out from other nucleic acid sequences. A "segment" is defined as a region of nucleic acid within the target sequence.

As used herein, the term "polymerase chain reaction" ("PCR") refers to the method of Mullis (*See e.g.*, U.S. Patent Nos. 4,683,195 4,683,202, and 4,965,188, hereby incorporated by reference), which describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing and polymerase extension can be repeated many times (*i.e.*, denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified".

With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (e.g., hybridization with a labeled

probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of <sup>32</sup>P-labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any oligonucleotide or polynucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications.

As used herein, the terms "PCR product," "PCR fragment," and "amplification product" refer to the resultant mixture of compounds after two or more cycles of the PCR steps of denaturation, annealing and extension are complete. These terms encompass the case where there has been amplification of one or more segments of one or more target sequences.

As used herein, the term "amplification reagents" refers to those reagents (deoxyribonucleotide triphosphates, buffer, etc.), needed for amplification except for primers, nucleic acid template and the amplification enzyme. Typically, amplification reagents along with other reaction components are placed and contained in a reaction vessel (test tube, microwell, etc.).

As used herein, the term "RT-PCR" refers to the replication and amplification of RNA sequences. In this method, reverse transcription is coupled to PCR, most often using a single enzyme procedure in which a thermostable polymerase is employed, as described in U.S. Patent No. 5,322,770, herein incorporated by reference. In RT-PCR, the RNA template is converted to cDNA due to the reverse transcriptase activity of the polymerase, and then amplified using the polymerizing activity of the polymerase (*i.e.*, as in other PCR methods).

As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence.

As used herein, the term "recombinant DNA molecule" as used herein refers to a DNA molecule comprising segments of DNA joined together by means of molecular biological techniques.

As used herein, the term "antisense" is used in reference to RNA sequences that are complementary to a specific RNA sequence (*e.g.*, mRNA). Included within this definition are antisense RNA ("asRNA") molecules involved in gene regulation by bacteria. Antisense RNA may be produced by any method, including synthesis by splicing the gene(s) of interest in a

reverse orientation to a viral promoter, permitting the synthesis of a coding strand. Once introduced into an embryo, this transcribed strand combines with natural mRNA produced by the embryo to form duplexes. These duplexes then block either the further transcription of the mRNA or its translation. In this manner, mutant phenotypes may be generated. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. The designation (-) (*i.e.*, "negative") is sometimes used in reference to the antisense strand, with the designation (+) sometimes used in reference to the sense (*i.e.*, "positive") strand.

As used herein, the term "polyA<sup>+</sup> RNA" refers to RNA molecules having a stretch of adenine nucleotides at the 3' end. This polyadenine stretch is also referred to as a "poly-A tail." Eukaryotic mRNA molecules contain poly-A tails and are referred to as polyA<sup>+</sup> RNA.

The terms "in operable combination," "in operable order," and "operably linked" as used herein refer to the linkage of nucleic acid sequences in such a manner that a nucleic acid molecule capable of directing the transcription of a given gene and/or the synthesis of a desired protein molecule is produced. The term also refers to the linkage of amino acid sequences in such a manner so that a functional protein is produced.

The term "isolated" when used in relation to a nucleic acid, as in "an isolated oligonucleotide" or "isolated polynucleotide" refers to a nucleic acid sequence that is identified and separated from at least one contaminant nucleic acid with which it is ordinarily associated in its natural source. Isolated nucleic acid is such present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids as nucleic acids such as DNA and RNA found in the state they exist in nature. For example, a given DNA sequence (*e.g.*, a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs, which encode a multitude of proteins. However, isolated nucleic acid encoding a mammalian  $\alpha 7$  protein includes, by way of example, such nucleic acid in cells ordinarily expressing an  $\alpha 7$  protein where the nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid, oligonucleotide, or polynucleotide may be present in single-stranded or double-stranded form. When an isolated nucleic acid, oligonucleotide or polynucleotide is to be utilized to express a protein, the

oligonucleotide or polynucleotide will contain at a minimum the sense or coding strand (*i.e.*, the oligonucleotide or polynucleotide may single-stranded), but may contain both the sense and anti-sense strands (*i.e.*, the oligonucleotide or polynucleotide may be double-stranded).

As used herein, a "portion of a chromosome" refers to a discrete section of the chromosome. Chromosomes are divided into sites or sections by cytogeneticists as follows: the short (relative to the centromere) arm of a chromosome is termed the "p" arm; the long arm is termed the "q" arm. Each arm is then divided into 2 regions termed region 1 and region 2 (region 1 is closest to the centromere). Each region is further divided into bands. The bands may be further divided into sub-bands. For example, the 11p15.5 portion of human chromosome 11 is the portion located on chromosome 11 (11) on the short arm (p) in the first region (1) in the 5th band (5) in sub-band 5 (.5). A portion of a chromosome may be "altered;" for instance the entire portion may be absent due to a deletion or may be rearranged (*e.g.*, inversions, translocations, expanded or contracted due to changes in repeat regions). In the case of a deletion, an attempt to hybridize (*i.e.*, specifically bind) a probe homologous to a particular portion of a chromosome could result in a negative result (*i.e.*, the probe could not bind to the sample containing genetic material suspected of containing the missing portion of the chromosome). Thus, hybridization of a probe homologous to a particular portion of a chromosome may be used to detect alterations in a portion of a chromosome.

The term "sequences associated with a chromosome" means preparations of chromosomes (*e.g.*, spreads of metaphase chromosomes), nucleic acid extracted from a sample containing chromosomal DNA (*e.g.*, preparations of genomic DNA); the RNA which is produced by transcription of genes located on a chromosome (*e.g.*, hnRNA and mRNA) and cDNA copies of the RNA transcribed from the DNA located on a chromosome. Sequences associated with a chromosome may be detected by numerous techniques including probing of Southern and Northern blots and *in situ* hybridization to RNA, DNA or metaphase chromosomes with probes containing sequences homologous to the nucleic acids in the above listed preparations.

As used herein the term "coding region" when used in reference to structural gene refers to the nucleotide sequences which encode the amino acids found in the nascent polypeptide as a result of translation of a mRNA molecule. The coding region is bounded, in eukaryotes, on the

5' side by the nucleotide triplet "ATG" which encodes the initiator methionine and on the 3' side by one of the three triplets that specify stop codons (*i.e.*, TAA, TAG, TGA).

As used herein, the term "structural gene" refers to a DNA sequence coding for RNA or a protein. In contrast, "regulatory genes" are structural genes that encode products that control the expression of other genes (*e.g.*, transcription factors).

As used herein, the term "purified" or "to purify" refers to the removal of contaminants from a sample. For example, anti- $\alpha 7$  antibodies are purified by removal of contaminating non-immunoglobulin proteins; they are also purified by the removal of immunoglobulin that does not bind  $\alpha 7$ . The removal of non-immunoglobulin proteins and/or the removal of immunoglobulins that do not bind  $\alpha 7$ , results in an increase in the percent of  $\alpha 7$ -reactive immunoglobulins in the sample. In another example, recombinant  $\alpha 7$  polypeptides are expressed in bacterial host cells and the polypeptides are purified by the removal of host cell proteins; the percent of recombinant  $\alpha 7$  polypeptides is thereby increased in the sample.

The term "recombinant DNA molecule" as used herein refers to a DNA molecule that is comprised of segments of DNA joined together by means of molecular biological techniques.

The term "recombinant protein" or "recombinant polypeptide" as used herein refers to a protein molecule that is expressed from a recombinant DNA molecule.

The term "native protein" as used herein to indicate that a protein does not contain amino acid residues encoded by vector sequences; that is the native protein contains only those amino acids found in the protein as it occurs in nature. A native protein may be produced by recombinant means or may be isolated from a naturally occurring source.

As used herein the term "portion" when in reference to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from four amino acid residues to the entire amino acid sequence minus one amino acid.

As used herein, the term "fusion protein" refers to a chimeric protein containing the protein of interest (*i.e.*, mouse or human  $\alpha 7$  and fragments thereof) joined to an exogenous protein fragment (the fusion partner which consists of a non- $\alpha 7$  protein). The fusion partner may enhance solubility of the  $\alpha 7$  protein as expressed in a host cell, may provide an affinity tag to allow purification of the recombinant fusion protein from the host cell or culture supernatant, or both. If desired, the fusion protein may be removed from the protein of interest (*i.e.*,  $\alpha 7$  protein or fragments thereof) by a variety of enzymatic or chemical means known to the art.

The term "Southern blot," refers to the analysis of DNA on agarose or acrylamide gels to fractionate the DNA according to size followed by transfer of the DNA from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized DNA is then probed with a labeled probe to detect DNA species complementary to the probe used. The DNA may be  
5 cleaved with restriction enzymes prior to electrophoresis. Following electrophoresis, the DNA may be partially depurinated and denatured prior to or during transfer to the solid support. Southern blots are a standard tool of molecular biologists (Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, NY, pp 9.31-9.58, 1989).

The term "Northern blot," as used herein refers to the analysis of RNA by electrophoresis  
10 of RNA on agarose gels to fractionate the RNA according to size followed by transfer of the RNA from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized RNA is then probed with a labeled probe to detect RNA species complementary to the probe used. Northern blots are a standard tool of molecular biologists (Sambrook *et al.*, *supra*, pp 7.39-7.52, 1989).

The term "Western blot" refers to the analysis of protein(s) (or polypeptides) immobilized  
15 onto a support such as nitrocellulose or a membrane. The proteins are run on acrylamide gels to separate the proteins, followed by transfer of the protein from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized proteins are then exposed to antibodies with reactivity against an antigen of interest. The binding of the antibodies may be detected by  
20 various methods, including the use of radiolabeled antibodies.

The term "antigenic determinant" as used herein refers to that portion of an antigen that makes contact with a particular antibody (*i.e.*, an epitope). When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure  
25 on the protein; these regions or structures are referred to as antigenic determinants. An antigenic determinant may compete with the intact antigen (*i.e.*, the "immunogen" used to elicit the immune response) for binding to an antibody.

The terms "specific binding" or specifically binding" when used in reference to the interaction of an antibody and a protein or peptide means that the interaction is dependent upon  
30 the presence of a particular structure (*i.e.*, the antigenic determinant or epitope) on the protein; in other words the antibody is recognizing and binding to a specific protein structure rather than to

proteins in general. For example, if an antibody is specific for epitope "A," the presence of a protein containing epitope A (or free, unlabelled A) in a reaction containing labeled "A" and the antibody will reduce the amount of labeled A bound to the antibody.

The present invention also contemplates "non-human animals" comprising any non-human animal capable of overexpressing  $\alpha 7$  mRNA and/or proteins. Such non-human animals include vertebrates such as rodents, non-human primates, ovines, bovines, ruminants, lagomorphs, porcines, caprines, equines, canines, felines, aves, etc. Preferred non-human animals are selected from the order Rodentia, most preferably mice. The term "order Rodentia" refers to rodents (*i.e.*, placental mammals [Class Euthria] which include the family Muridae (rats and mice).

The "non-human animals having a genetically engineered genotype" of the invention are preferably produced by experimental manipulation of the genome of the germline of the non-human animal. These genetically engineered non-human animals may be produced by several methods including the introduction of a "transgene" comprising nucleic acid (usually DNA) into an embryonal target cell or integration into a chromosome of the somatic and/or germ line cells of a non-human animal by way of human intervention, such as by the methods described herein. Non-human animals that contain a transgene are referred to as "transgenic non-human animals." A transgenic animal is an animal whose genome has been altered by the introduction of a transgene.

The term "transgene" as used herein refers to a foreign gene that is placed into an organism by introducing the foreign gene into newly fertilized eggs or early embryos. The term "foreign gene" refers to any nucleic acid (*e.g.*, gene sequence) that is introduced into the genome of an animal by experimental manipulations and may include gene sequences found in that animal so long as the introduced gene does not reside in the same location as does the naturally occurring gene.

As used herein, the term "vector" is used in reference to nucleic acid molecules that transfer DNA segment(s) from one cell to another. The term "vehicle" is sometimes used interchangeably with "vector."

The term "expression vector" as used herein refers to a recombinant DNA molecule containing a desired coding sequence and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host organism. Nucleic acid

sequences necessary for expression in prokaryotes usually include a promoter, an operator (optional), and a ribosome-binding site, often along with other sequences. Eukaryotic cells are known to utilize promoters, enhancers, and termination and polyadenylation signals.

The terms "overexpression" and "overexpressing" and grammatical equivalents, are used in reference to levels of mRNA to indicate a level of expression approximately 3-fold higher than that typically observed in a given tissue in a control or non-transgenic animal. Levels of mRNA are measured using any of a number of techniques known to those skilled in the art including, but not limited to Northern blot analysis. Appropriate controls are included on the Northern blot to control for differences in the amount of RNA loaded from each tissue analyzed (*e.g.*, the amount of 28S rRNA, an abundant RNA transcript present at essentially the same amount in all tissues, present in each sample can be used as a means of normalizing or standardizing the  $\alpha 7$  mRNA-specific signal observed on Northern blots). The amount of mRNA present in the band corresponding in size to the correctly spliced  $\alpha 7$  transgene RNA is quantified; other minor species of RNA which hybridize to the transgene probe are not considered in the quantification of the expression of the transgenic mRNA.

The term "transfection" as used herein refers to the introduction of foreign DNA into eukaryotic cells. Transfection may be accomplished by a variety of means known to the art including calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection, polybrene-mediated transfection, electroporation, microinjection, liposome fusion, lipofection, protoplast fusion, retroviral infection, and biolistics.

The term "stable transfection" or "stably transfected" refers to the introduction and integration of foreign DNA into the genome of the transfected cell. The term "stable transfectant" refers to a cell, which has stably integrated foreign DNA into the genomic DNA.

The term "transient transfection" or "transiently transfected" refers to the introduction of foreign DNA into a cell where the foreign DNA fails to integrate into the genome of the transfected cell. The foreign DNA persists in the nucleus of the transfected cell for several days. During this time the foreign DNA is subject to the regulatory controls that govern the expression of endogenous genes in the chromosomes. The term "transient transfectant" refers to cells that have taken up foreign DNA but have failed to integrate this DNA.

The term "calcium phosphate co-precipitation" refers to a technique for the introduction of nucleic acids into a cell. The uptake of nucleic acids by cells is enhanced when the nucleic

acid is presented as a calcium phosphate-nucleic acid co-precipitate. The original technique of Graham and van der Eb (Graham and van der Eb, *Virology*, 52:456, 1973), has been modified by several groups to optimize conditions for particular types of cells. The art is well aware of these numerous modifications.

5 As used herein, the term "selectable marker" refers to the use of a gene which encodes an enzymatic activity that confers the ability to grow in medium lacking what would otherwise be an essential nutrient (*e.g.*, the *HIS3* gene in yeast cells); in addition, a selectable marker may confer resistance to an antibiotic or drug upon the cell in which the selectable marker is expressed. Selectable markers may be "dominant"; a dominant selectable marker encodes an enzymatic activity, which can be detected in any eukaryotic cell line. Examples of dominant  
10 selectable markers include the bacterial aminoglycoside 3' phosphotransferase gene (also referred to as the *neo* gene) which confers resistance to the drug G418 in mammalian cells, the bacterial hygromycin G phosphotransferase (*hyg*) gene which confers resistance to the antibiotic hygromycin and the bacterial xanthine-guanine phosphoribosyl transferase gene (also referred to  
15 as the *gpt* gene) which confers the ability to grow in the presence of mycophenolic acid. Other selectable markers are not dominant in that their use must be in conjunction with a cell line that lacks the relevant enzyme activity. Examples of non-dominant selectable markers include the thymidine kinase (*tk*) gene, which is used in conjunction with *tk*<sup>-</sup> cell lines, the CAD gene, which is used in conjunction with CAD-deficient cells and the mammalian hypoxanthine-guanine  
20 phosphoribosyl transferase (*hprt*) gene, which is used in conjunction with *hprt*<sup>-</sup> cell lines. A review of the use of selectable markers in mammalian cell lines is provided in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, New York, pp.16.9-16.15, 1989.

25 As used herein, the term "cell culture" refers to any *in vitro* culture of cells. Included within this term are continuous cell lines (*e.g.*, with an immortal phenotype), primary cell cultures, finite cell lines (*e.g.*, non-transformed cells), and any other cell population maintained *in vitro*.

The term "compound" refers to any chemical entity, pharmaceutical, drug, and the like that can be used to treat or prevent a disease, illness, sickness, or disorder of bodily function.  
30 Compounds comprise both known and potential therapeutic compounds. A compound can be determined to be therapeutic by screening using the screening methods of the present invention.

A "known therapeutic compound" refers to a therapeutic compound that has been shown (*e.g.*, through animal trials or prior experience with administration to humans) to be effective in such treatment. In other words, a known therapeutic compound is not limited to a compound efficacious in the treatment of cancer.

5           The term "test compound" refers to any chemical entity, pharmaceutical, drug, and the like that can be used to treat or prevent a disease, illness, sickness, or disorder of bodily function. Test compounds comprise both known and potential therapeutic compounds. A test compound can be determined to be therapeutic by screening using the methods of the present invention. A "known therapeutic compound" refers to a therapeutic compound that has been shown (*e.g.*,  
10 through animal trials or prior experience with administration to humans) to be effective in such treatment or prevention. In other words, a known therapeutic compound is not limited to a compound efficacious in the treatment of  $\alpha 7$  instability or inactivity in animals.

A "composition comprising a given polynucleotide sequence" as used herein refers broadly to any composition containing the given polynucleotide sequence. The composition may  
15 comprise an aqueous solution. Compositions comprising polynucleotide sequences encoding human  $\alpha 7$  (*e.g.*, SEQ ID NO:123), or fragments thereof, may be employed as hybridization probes. In other embodiments, compositions comprising the promoter and upstream untranslated sequence of human  $\alpha 7$  (*e.g.*, SEQ ID NO:122) or fragments thereof (*e.g.*, SEQ ID NO:94, 101, 125, etc) may be employed as hybridization probes. In these cases, the human  $\alpha 7$ -encoding  
20 polynucleotide sequences are typically employed in an aqueous solution containing salts (*e.g.*, NaCl), detergents (*e.g.*, SDS) and other components (*e.g.*, Denhardt's solution, dry milk, salmon sperm DNA, etc.).

The term "sample" as used herein is used in its broadest sense. A sample suspected of containing a human chromosome or sequences associated with a human chromosome may  
25 comprise a cell, chromosomes isolated from a cell (*e.g.*, a spread of metaphase chromosomes), genomic DNA (in solution or bound to a solid support such as for Southern blot analysis), RNA (in solution or bound to a solid support such as for Northern blot analysis), cDNA (in solution or bound to a solid support) and the like. A sample suspected of containing a protein may comprise a cell, a portion of a tissue, an extract containing one or more proteins and the like.

30           The term "test  $\alpha 7$ " refers to a sample suspected of containing  $\alpha 7$ . The concentration of  $\alpha 7$  in the test sample is determined by various means, and may be compared with a "quantitated

amount of  $\alpha 7$ " (*i.e.*, a positive control sample containing a known amount of  $\alpha 7$ ), in order to determine whether the concentration of test  $\alpha 7$  in the sample is within the range usually found within samples from wild-type organisms. Thus, comparison of the positive control with the test sample allows the determination to be made whether a particular individual produces a "normal" amount of  $\alpha 7$ , is deficient in production of  $\alpha 7$ , or produces a concentration of  $\alpha 7$  that is greater than normal. It is intended that such test methods also contain "negative" controls (*i.e.*, samples that are known to contain no  $\alpha 7$ ). Furthermore, it is intended that the testing be conducted using the  $\alpha 7$  gene,  $\alpha 7$  mRNA, and/or  $\alpha 7$  protein (or polypeptides), or fragments of any of these.

The term "heteroduplex analysis" as used herein refers to a method of detecting mutations based on the retardation of the heteroduplex compared with the corresponding homoduplex on a non-denaturing polyacrylamide gel. Heteroduplexes migrate more slowly than their corresponding homoduplexes due to a more open double-stranded configuration surrounding the mismatched bases.

As used herein, the terms "DHPLC" and "denaturing high performance liquid chromatography" refer to a scanning method for mutation detection based on the capability of ion-pair reverse phase liquid chromatography on alkylated nonporous particles to resolve homo from heteroduplex molecules under conditions of partial denaturation (Underhill *et al.*, *Proc Natl Acad Sci USA*, 93:196-2000, 1996 and U.S. Patent Nos. 5,795,976, herein incorporated by reference in their entirety).

The terms "single-strand conformation polymorphism" and "SSCP," as used herein, refer to the ability of single strands of nucleic acid to take on characteristic conformations under non-denaturing conditions, which in turn can influence the electrophoretic mobility of the single-stranded nucleic acids. Changes in the sequence of a given fragment (*i.e.*, mutations) will also change the conformation, consequently altering the mobility and allowing this to be used as an assay for sequence variations (Orita *et al.*, *Genomics* 5:874-879, 1989).

As used herein, the terms "conformation-sensitive gel electrophoresis" or "CSGE" refer to methods for detecting mutations involving distinguishing DNA heteroduplexes from homoduplexes via mildly denaturing gel electrophoresis. CSGE protocols are well known in the art (Ganguly *et al.*, *Proc Natl Acad Sci USA* 90:10325-10329, 1993).

As used herein, the terms "ligase chain reaction" and "ligase amplification reaction" refer to methods for detecting small quantities of a target DNA, with utility similar to PCR. Ligase chain reaction relies on DNA ligase to join adjacent synthetic oligonucleotides after they have bound the target DNA. Their small size means that they are destabilized by single base mismatches and so form a sensitive test for the presence of mutations in the target sequence.

The term "DNA sequencing" refers to methods used to determine the order of nucleotide bases in a DNA molecule or fragment. The term "DNA sequencing" includes for example, dideoxy sequencing and Maxam-Gilbert sequencing.

As used herein, the term "*in vitro*" refers to an artificial environment and to processes or reactions that occur within an artificial environment. In vitro environments can consist of, but are not limited to, test tubes and cell culture. The term "*in vivo*" refers to the natural environment (*e.g.*, an animal or a cell) and to processes or reaction that occur within a natural environment.

The terms "test compound" and "candidate compound" refer to any chemical entity, pharmaceutical, drug, and the like that is a candidate for use to treat or prevent a disease, illness (*e.g.*, major depressive disorder), sickness, or disorder of bodily function. Test compounds comprise both known and potential therapeutic compounds. A test compound can be determined to be therapeutic by screening using the screening methods of the present invention.

The term "change" as used herein refers to a difference or a result of a modification or alteration. In preferred embodiments, the term "change" refers to a measurable difference between states (*e.g.*, higher or lower  $\alpha 7$  mRNA or protein expression in a cell in the presence and absence of a test compound). In some embodiments, the change is at least 10%, preferably at least 25%, more preferably at least 50%, and most preferably at least 90% more or less than that of a control condition.

As used herein, the term "sample" is meant to include a specimen obtained from subject. The term "sample" encompasses fluids, solids, and tissues. In preferred embodiments, the term "sample" refers to blood or biopsy material obtained from a living body for the purpose of examination via any appropriate technique (*e.g.*, needle, sponge, scalpel, swab, *etc.*). In particularly preferred embodiments, the term "sample" refers to buccal cells (*e.g.*, cells of the inner lining of the mouth or cheek). Buccal cell samples are obtained using any suitable method,

including but not limited to collection via tongue depressor, cytobrush or mouthwash (*See, Moore et al., Biomarkers, 6:448-454, 2001*).

The terms "subject" as used herein, refers to a human. It is intended that the term encompass healthy individuals, as well as, individuals predisposed to, or suspected of having schizophrenia. Typically, the terms "subject" and "patient" are used interchangeably. In some preferred embodiments of the present invention, the term subject refers to specific subgroups of patients.

The term "schizophrenia" as used herein refers to a major mental disorder featuring psychotic symptoms during some phase of the illness, a long term course and a deterioration in function. Schizophrenic symptoms can be classified as positive, negative, cognitive and mood symptoms, which together or separately may result in behavioral disturbances (*e.g.*, bizarre, apparently purposeless and stereotyped activity or inactivity). Various embodiments of the present invention are contemplated to effectively treat all subtypes of schizophrenia, including but not limited to catatonic, disorganized, paranoid and undifferentiated subtypes. In addition, the compositions and methods of the present invention are also contemplated to benefit patients with schizoid personality disorder (socially distant, detached) and patients with schizotypal personality disorder (odd, eccentric).

As used herein, the term "positive symptoms" refers to symptoms including but not limited to hallucinations (*e.g.*, hearing voices), delusions (*e.g.*, of persecution or grandiosity), disorganized speech and thought, altered sense of self and bizarre behavior. They are called positive symptoms because they are added on to the individuals experience and behavior.

The term "negative symptoms" as used herein, refers to deficit symptoms, including experience and behavior that should be there and is not. Negative symptoms include but are not limited to loss of motivation, flattened emotions, withdrawal from an active social life, poverty of thought and speech, and loss of former interests and pleasures.

As used herein, the term "cognitive symptoms," refers to symptoms associated with a loss of cognitive ability including but not limited to attention deficits, memory loss, inability to plan for the future and poor capacity for abstract thought.

The term "mood symptoms" as used herein, refers to symptoms associated with a disturbed state of mind or predominant emotion such as dysphoria.

As used herein, the term "risk of developing schizophrenia" refers to a subject's relative risk (*e.g.*, the percent chance or a relative score) of developing schizophrenia during their lifetime.

The term "subject suspected of having schizophrenia" refers to a subject that presents one or more symptoms indicative of schizophrenia (*e.g.*, delusions, hallucinations, disorganized speech, catatonic behavior, negative symptoms such as effective flattening, alogia or avolition, etc.) or is being screened for schizophrenia (*e.g.*, during a routine physical).

As used herein, the term "diagnosis" refers to the determination of the nature of a case of disease. In some preferred embodiments of the present invention, methods for making a diagnosis are provided which permit schizophrenia to be distinguished from other forms of mental illness including but not limited to psychosis due to a general medical condition; delirium, or dementia; substance-induced or related disorders; depressive disorder; and bipolar disorder (*e.g.*, manic depression).

The term "reagent(s) suitable for use in specifically detecting at least one polymorphism in an  $\alpha 7$  allele" refers to reagent(s) used to detect a polymorphism of interest in an  $\alpha 7$  gene, cDNA, or RNA. Examples of suitable reagents include but are not limited to, nucleic acid probes and primers capable of specifically hybridizing to  $\alpha 7$  mRNA or cDNA. In some preferred embodiments, the term suitable reagents refers to primers for amplifying an  $\alpha 7$  fragment suspected of containing a polymorphism of interest.

As used herein, the term "instructions for determining whether a subject is predisposed to schizophrenia" refers to instructions for using the reagents contained in the kit for the detection and characterization of an  $\alpha 7$  allele in a sample from a subject. In some embodiments, the instructions further comprise the statement of intended use required by the U.S. Food and Drug Administration (FDA) in labeling *in vitro* diagnostic products. The FDA classifies *in vitro* diagnostics as medical devices and required that they be approved through the 510(k) procedure. Information required in an application under 510(k) includes: 1) The *in vitro* diagnostic product name, including the trade or proprietary name, the common or usual name, and the classification name of the device; 2) The intended use of the product; 3) The establishment registration number, if applicable, of the owner or operator submitting the 510(k) submission; the class in which the *in vitro* diagnostic product was placed under section 513 of the FD&C Act, if known, its appropriate panel, or, if the owner or operator determines that the device has not been

classified under such section, a statement of that determination and the basis for the determination that the in vitro diagnostic product is not so classified; 4) Proposed labels, labeling and advertisements sufficient to describe the in vitro diagnostic product, its intended use, and directions for use, including photographs or engineering drawings, where applicable; 5) A statement indicating that the device is similar to and/or different from other in vitro diagnostic products of comparable type in commercial distribution in the U.S., accompanied by data to support the statement; 6) A 510(k) summary of the safety and effectiveness data upon which the substantial equivalence determination is based; or a statement that the 510(k) safety and effectiveness information supporting the FDA finding of substantial equivalence will be made available to any person within 30 days of a written request; 7) A statement that the submitter believes, to the best of their knowledge, that all data and information submitted in the premarket notification are truthful and accurate and that no material fact has been omitted; and 8) Any additional information regarding the in vitro diagnostic product requested that is necessary for the FDA to make a substantial equivalency determination. Additional information is available at the Internet web page of the U.S. FDA.

## EXPERIMENTAL

The following examples are provided in order to demonstrate and further illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

5           In the experimental disclosure which follows, the following abbreviations apply: nAChR (nicotinic acetylcholine receptor); °C (degrees Centigrade); rpm (revolutions per minute); BSA (bovine serum albumin); H<sub>2</sub>O (water); HCl (hydrochloric acid); aa (amino acid); bp (base pair); kb or kbp (kilobase pair); Mb (megabase pair); kD (kilodaltons); gm or g (grams); µg (micrograms); mg (milligrams); ng (nanograms); µl (microliters); ml (milliliters); mm  
10 (millimeters); nm (nanometers); µm (micrometer); M (molar); mM (millimolar); µM (micromolar); nM (nanomolar); pM (picomolar); U or u (units); V (volts); MW (molecular weight); sec (seconds); min(s) (minute/minutes); hr(s) (hour/hours); dNTP (deoxynucleotide); MgCl<sub>2</sub> (magnesium chloride); NaCl (sodium chloride); KCl (potassium chloride); DTT (dithiotreitol); DMSO (dimethyl sulfoxide); NaOH (sodium hydroxide); 3'UT (3'-untranslated  
15 region); OD<sub>280</sub> (optical density at 280 nm); OD<sub>600</sub> (optical density at 600 nm); EST (expressed sequence tag); PAGE (polyacrylamide gel electrophoresis); PBS (phosphate buffered saline [150 mM NaCl, 10 mM sodium phosphate buffer, pH 7.2]); PCR (polymerase chain reaction); DMEM (Dulbecco's Modified Eagle Medium); PEG (polyethylene glycol); PMSF (phenylmethylsulfonyl fluoride); RT-PCR (reverse transcription PCR); SDS (sodium dodecyl sulfate); SSC (saline-  
20 sodium citrate buffer); Tris (tris(hydroxymethyl)aminomethane); w/v (weight to volume); v/v (volume to volume); YAC (yeast artificial chromosome); BAC (bacterial artificial chromosome); PAC (P1 artificial chromosome); RACE (Rapid Amplification of cDNA Ends); TAFE (Transverse Alternating Field Electrophoresis); lod (maximum logarithm of the odds); STS (sequence-tagged site); Beckman (Beckman Instruments, Inc., Fullerton, CA); Amersham  
25 (Amersham Life Science, Inc. Arlington Heights, IL); Qiagen (Qiagen Inc., Santa Clarita, CA); Genome Systems (Genome Systems, St. Louis, MO, USA); ICN (ICN Pharmaceuticals, Inc., Costa Mesa, CA); Amicon (Amicon, Inc., Beverly, MA); NCBI (National Center for Biotechnology Information, Bethesda, MD); ATCC (American Type Culture Collection, Rockville, MD); Research Genetics (Research Genetics, Huntsville, AL); Pharmacia  
30 (Pharmacia and Upjohn Diagnostics, Kalamazoo, MI); Boehringer-Mannheim (Boehringer-Mannheim, Indianapolis, IN); National Biosciences (National Biosciences, Inc., Plymouth MN);

MJ Research (MJ Research, Watertown, MA); Perkin-Elmer (Perkin-Elmer, Foster City, CA); BioRad (BioRad, Richmond, CA); Clontech (CLONTECH Laboratories, Palo Alto, CA); Gibco, GIBCO BRL, or Gibco BRL (Life Technologies, Inc., Gaithersburg, MD); Gene Codes (Gene Codes Corporation, Ann Arbor, MI); Invitrogen (Invitrogen Corp., San Deigo, CA); Kodak (Eastman Kodak Co., New Haven, CT); Promega (Promega, Corp., Madison, WI); New England Biolabs (New England Biolabs, Inc., Beverly, MA); Novagen (Novagen, Inc., Madison, WI); Pharmacia (Pharmacia, Inc., Piscataway, NJ); Schleicher & Schuell (Schleicher and Schuell, Inc., Keene, NH); Sigma (Sigma Chemical Co., St. Louis, MO); Sorvall (Sorvall Instruments, a subsidiary of DuPont Co., Biotechnology Systems, Wilmington, DE); Stratagene (Stratagene Cloning Systems, La Jolla, CA); Whatman (Whatman LabSales, Hillsboro, OR); Bethyl Laboratories (Bethyl Laboratories, Montgomery, TX); Ambion (Ambion, Inc., Austin, TX); and Zeiss (Carl Zeiss, Inc., Thornwood, NY).

Unless otherwise indicated, all restriction enzymes were obtained from New England BioLabs and were used according to the manufacturer's instructions.

## EXAMPLE 1

### Samples

Samples were obtained from various normal individuals for use as controls in the Examples described below. To prepare these samples, blood was drawn from normal subjects, seen in the Denver Schizophrenia Center. Of the 43 subjects used for the polymorphism analysis, 22 were female and 21 were male. There were 38 Caucasians, 2 Blacks, 1 Asian and 2 Hispanics. None of the subjects had a history of mental illness or a family history of mental illness.

In addition to the "normal" samples, pedigrees were selected for presence of at least two cases of schizophrenia in a nuclear family. Two psychiatrists made clinical diagnoses of schizophrenia, chronic type, blind to pedigree and genetic information, using Research Diagnostic Criteria (Spitzer *et al.*, Arch Gen Psychiat, 35:773, 1978; and Endicott and Spitzer, Arch Gen Psychiat, 35:837, 1978). Nine families with 104 members were studied (*i.e.*, nine pedigrees containing individuals diagnosed with schizophrenia were analyzed). All subjects gave written informed consent. Blood was also drawn from these individuals for DNA analysis.

DNA was extracted from blood samples as described by Miller *et al.* (Miller *et al.*, Nuc Acids Res, 16:1215, 1988), with one additional step. Briefly, red blood cells were lysed by incubating 10-15 ml of anticoagulated blood at 4 C for 10 minutes in 40 ml blood cell lysis solution (BCL) (BCL contains 0.3 M sucrose, 0.01 M Tris HCl pH 7.5, 0.005 M MgCl<sub>2</sub> and 1% Triton X-100) with occasional rocking to mix. The cells were then centrifuged at 850xg at 4 C for 15 minutes. The pellet was resuspended by repeated pipetting with a 1 ml wide bore glass pipet in 30 ml BCL (4°C) and centrifuged as before.

DNA was then extracted from the pellet as described (Miller *et al.*, Nucl Acids Res, 16:1215, 1988). Briefly, the pellet was resuspended as before in 3 ml Nuclei Lysis buffer (NL) (NL contains 0.075 M NaCl, and 0.024 M EDTA pH 8.0). Then, 200 µl of 10% SDS, 440 µl of digest diluent (1% SDS, 2 mM Na<sub>2</sub>EDTA), and 60 µl of Proteinase K (20 mg/ml stock) were then added to the suspension. The suspension was then incubated at 37°C for 16-20 hours with gentle mixing by inversion. Following this digestion, 1 ml of saturated (approx. 6 M) NaCl was added and then the suspension briefly (15 seconds) was vigorously shaken. The suspension was then centrifuged at 1340xg at room temperature for 15 minutes. The supernatant was transferred to a new tube, leaving the pellet at the bottom of the previous tube undisturbed. Exactly 2 volumes of absolute ethanol were added. The tube was then inverted several times until the DNA pellet was visible and floated to the top. The pellet was then transferred to a new tube. The pellet was resuspended in 0.67 ml TE pH 8 (10 mM Tris, 1 mM EDTA) by gently mixing on a roller drum for 3-5 days at 37°C.

In addition to the samples described above, a chromosome 15 somatic cell hybrid line, R379-2B2 generously provided by Dr. Carol Jones (The Eleanor Roosevelt Institute for Cancer Research, Denver, CO), was also used. This cell line was cultured in Ham's F12, supplemented with 5% fetal bovine serum and 10 µg/ml gentamicin.

Another cell line, the human neuroblastoma cell line, SH-SY5Y (Biedler *et al.*, Cancer Res, 38:3751, 1978), was obtained from Dr. June Biedler (Memorial Sloan-Kettering Cancer Center, New York, NY), and grown in DMEM/Ham's F12 (1:1 ratio, supplemented with 15% fetal bovine serum, 4 mM glutamine, and 10 µg/ml gentamicin).

## EXAMPLE 2

### Genomic Clone Isolation

In this Example, YAC clones were identified by PCR screening of two genomic libraries, namely the St. Louis YAC library (Burke *et al.*, Science, 236:806, 1987) and the CEPH YAC Library 3 (Albertsen *et al.*, Proc Natl Acad Sci, USA, 87:4256, 1990), using  $\alpha 7$  cDNA specific primers and methods known in the art (*See e.g.*, Brownstein *et al.*, Science, 244:1348, 1989; Chumakov *et al.*, Nature, 359:380, 1992; and Dracopoli *et al.*, *Current Protocols in Human Genetics*, John Wiley & Sons, Inc., New York, NY, 1994).

Additional YACs, positive for loci in the  $\alpha 7$  nAChR region were identified initially by using Infocloner on the CEPH/Genethon Integrated Map courtesy of the Fondation Jean Dausset - CEPH world wide web server. Loci on the YAC contig were verified by PCR screening with either  $\alpha 7$  primer sets or primer sets for the specific polymorphic markers listed in the YAC contig (*See*, Figure 6), which are available from either the CEPH database or GenBank. The PCR conditions were 94°C for 2 minutes, 1 cycle; followed by 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and then 72°C for 30 seconds, followed by 72°C for 7 minutes-1 cycle. These PCR conditions were used for all PCR amplifications, unless otherwise indicated.

In addition,  $\alpha 7$ -specific primer sets were used to identify the two PAC clones 64a1 and 25919. A Research Genetics BAC library was screened with  $\alpha 7$  coding region primers by PCR to identify the BAC clone 467o18. The BAC library purchased from Research Genetics was a “pooled DNA” library, with each hit-positive PCR product of correct size being indicative of a location on a subsequent plate of pools. A hit on this plate gave an address to yet another plate, where the clone of interest was found. These “BAC clone” plates are maintained by Research Genetics. When the positive PCR reactions produced a final plate address in the clone library, that clone was ordered from Research Genetics. The PCR conditions and primers were as described herein (the primers used are shown in Tables 1 and 2). The two PAC clones (64a1 and 25919) were identified using the following PCR primers.

The primers used in these experiments were:

TCCTGATGTCGGCTCCCAACT (SEQ ID NO:1)

GGTACGGATGTGCCAAGGATA (SEQ ID NO:2)

TTTGGGGGTGCTAATCCAGGA (SEQ ID NO:3)

TTGTTTTCTTCCACCAGTCA (SEQ ID NO:4)

CTCGCTGCAGCTCCGGGACTCA (SEQ ID NO:5)

GGAGGCTCAGGGAGAAGTAG (SEQ ID NO:6)

The first two sets of primers were used to amplify the 3' untranslated region of the gene and the third primer set was used to amplify the first and second exons of the gene with the intervening intron 1 sequence. All PCR reactions were optimized in a Perkin Elmer 480 PCR using normal human DNA and cDNA. Conditions were as follows for the control reactions in the 3' sets: 96°C for two min, then 35 cycles of 96°C for 30 sec, 56°C for 30 sec, 72°C for 1 min, and cool to 4°C, using 4 mM MgCl<sub>2</sub> and 10% DMSO. The 5' PCR set was used in 1.5 mM MgCl<sub>2</sub> and 10%DMSO with the following conditions: 96°C for 3 min, then cycles 1-6 were 94°C for 1 min, 68°C -58°C for 1 min (dropping from 68°C to 58°C, by 2°C increments each cycle), 72° C for 1 min, followed by 30 cycles of 94°C for 1 min, 58°C for 1 min, 72°C for 1 min, then a 4 min extension at 4°C, followed by chilling at 4°C.

In these experiments, a genomic clone of the human  $\alpha 7$  nicotinic receptor was identified. A YAC designated as b134h10 of approximately 250 kb, was isolated from the St. Louis YAC library. A Southern blot comparison of YAC b134h10 with human genomic DNA indicated that it contained the full-length  $\alpha 7$  nAChR gene. This YAC was used to isolate a polymorphic marker, D15S1360, as described in greater detail below.

The polymorphic marker D15S1360, a complex microsatellite with four alleles, was isolated from a YAC containing the  $\alpha 7$  nicotinic receptor gene. The GenBank sequence for rat  $\alpha 7$  (GenBank Accession No. M85273) was used to design primers to the conserved regions of the  $\alpha 7$  coding sequence. These primers were then used to PCR amplify normal human hippocampal cDNA obtained from a normal brain postmortem. The products were sequenced by Automated dye-terminator chemistry (as described in Example 5). The human sequence in transmembrane regions III and IV was then used to design PCR primers. These primers were: 5'-CTCCAGGATC TTGGCCAAGT C-3' (sense strand; SEQ ID NO:7), and 5'-AGATGCCCAA GTGGACCAGA G-3' (antisense strand; SEQ ID NO:8).

The PCR reactions were conducted with 2 mM MgCl<sub>2</sub> and 10% DMSO, in a Perkin-Elmer 4800 using the following cycles: 94°C for 2 min, then 5 cycles of 94°C for 1 min, 56°C for 30 sec, 72°C for 1 min, then 35 cycles 94°C for 1 min, 54°C for 30 sec, 72°C for 1 min, and cooling at 4°C. The product was reamplified with primers extended to contain a sense *Xba* and antisense *Bam* site. The products were cut and ligated into a BlueScript SK- vector. Sequence of the probe was confirmed by automated dye-primer sequencing. Subsequent PCR based screening of the original YAC clones were based on the above primers and conditions, substituting YAC DNA for hippocampal cDNA as the template.

The PCR fragment (*i.e.*, the probe) was sequenced and human primers were designed to generate a 338 bp product, which was cloned into pBluescript SK(-). The sequence of the 338 bp probe was:

AGATGCCCAAGTGGACCAGAGTCATCCTTCTGAACTGGTGCGCGTGGTTCCTGCGAA  
TGAAGAGGCCCCGGGGAGGACAAGGTGCGCCCGGCCTGCCAGCACAAGCAGCGGCG  
CTGCAGCCTGGCCAGTGTGGAGATGAGCGCCGTGGGGCCCGCCGCCAGCAACG  
GGAACCTGCTGTACATCGGCTTCCGCGGCCTGGACGGCGTGCACTGTGTCCCGACCC  
CCGACTCTGGGGTAGTGTGTGGCCGCATGGCCTGCTCCCCACGCACGATGAGCACC  
TCCTGCACGGCGGGCAACCCCCCGAGGGGGACCCGGACTTGGCCAAGATCCTGGA  
(SEQ ID NO:9).

This probe was used to isolate a human  $\alpha 7$  cDNA (GenBank Accession No. U40583).

The Washington University human YAC library was screened with the same primers. Two clones were isolated, B132H10 (150 kbp) and B134H10 (300 kbp), on the TAFE (Beckman) gel system, using the procedures recommended by the manufacturer.

A sub-library of B134H10 was prepared in the  $\lambda$ ZAP phagemid vector by complete *Mbo*I digestion of the intact YAC DNA in a low-melt agarose plug. The DNA was extracted and ligated into *Bam*H1 digested and phosphatased vector, transformed into XL1Blue-(MRF'), and screened with a (CA)<sub>16</sub> (SEQ ID NO:10) oligonucleotide. One clone contained a microsatellite [(CA)<sub>5</sub>T(CA)<sub>12</sub>TA(CA)<sub>5</sub>C(CA)<sub>3</sub>] (SEQ ID NO:11), which mapped to chromosome 15 (Human/Rodent Hybrid Mapping Panel #1, Coriell, Camden NJ). Flanking primers amplified seven additional alleles (97, 107, 109, 111, 113, 115, and 117 bp). The primers used were 5'-GATCTTTGGT AGAAGC-3' (SEQ ID NO:12), and 5'-ACCACCACTA CCATACAGAC-3' (SEQ ID NO:13). Allele frequencies (0.006, 0.006, 0.006, 0.516, 0.370, 0.090, and 0.006;

heterozygosity 0.57) were estimated from individuals marrying into the pedigrees described in Example 1. Primer sets used for mapping  $\alpha 7$  exons to YAC clones are listed in Table 1, below. Primers used for mapping STS/dinucleotide repeat markers to YAC clones were obtained from GenBank, and are listed in Table 2. In these Tables, and unless otherwise indicated, all DNA sequences are shown in 5' to 3' orientation.

**Table 1. Primer Sets Used to Amplify Exon and Flanking Intron Sequence from Human Alpha-7 Nicotinic Acetylcholine Receptor**

<b>Sequence Amplified Primer Number</b>	<b>Sequence</b>	<b>SEQ ID NO:</b>
Promoter #1234	CAAAGAACGCAAGGGAGAGGT	SEQ ID NO:14
Promoter #1235	CGGCTCGCGCGCCTTTAAGGA	SEQ ID NO:15
Exon 1 #1331 or #1236	GGGCTCGTCACGTGGAAAAGC	SEQ ID NO:16
Exon 1 #1233	GGATCCCACGGAGGAGTGGAG	SEQ ID NO:17
Exon 2 #1198	CCTGCCCCGGGTCTTCTCTCCT	SEQ ID NO:18
Exon 2 #1138	AACTAGAGTGCCCCAGCCGAGCT	SEQ ID NO:19
Exon 3 #1475	AACAACGCTCTCGACAGTCAGATC	SEQ ID NO:20
Exon 3 #1476	AAGATCTTGCAGCCCATGGGAG	SEQ ID NO:21
Exon 4 #1368	GGAATTCTCTTTGGTTTTGCAC	SEQ ID NO:22
Exon 4 #1369	ACATATCCAGCATCTCTGTGA	SEQ ID NO:23
Exon 5 #1218	TCATGCAGTCCTTTTCCTGTTTC	SEQ ID NO:24
Exon 5 #1142	CTCGCTTCAGTTTCTAACATGG	SEQ ID NO:25
Exon 6 #1124	GGAAGTGTGTGTATTTTCAGC	SEQ ID NO:26
Exon 6 #1144	TTAAAGCTTGCCCAGGAATAGG	SEQ ID NO:27
Exon 7 #1143	GCTTGTGTGTGGTATACACATTG	SEQ ID NO:28
Exon 7 #1126	TCCAGAGCTGATCTCAGCAGAAG	SEQ ID NO:29
Exon 8 #1125	GCCCCTCGTTAGACAGAATTGAG	SEQ ID NO:30
Exon 8 #1145	CTGGGCACACTCTAACCTAACC	SEQ ID NO:31
Exon 9 #1146	TGTGACGTGCAGTGCCACAGGA	SEQ ID NO:32
Exon 9 #1127	AAAACCCTAGGAGGAGCCTCCTT	SEQ ID NO:33

Exon 10 #1128	GATCAGCCCGTTTCCGCCTCA	SEQ ID NO:34
Exon 10 #589	GGTACGGATGTGCCAAGGATA	SEQ ID NO:35
Exon A #1516	GGACTCTGCTTTTGATAAATATGTATG	SEQ ID NO:36
Exon A #1517	TTGCTGTCACTTTCTGTGTTTCAT	SEQ ID NO:37
Exon B #1283	GACAATCCAAAGGTGCAGAAAGC	SEQ ID NO:38
Exon B #1538	TTCGTATCTGTATACAGACAGTC	SEQ ID NO:39
Exon C #1567	CCTCAGCATCATATTAGTTCAGTG	SEQ ID NO:40
Exon C # 1572	GCGGACAAGAGAAACAGGAAAG	SEQ ID NO:41
Exon D #1534	GGCAGTGGTGCTGTTGCCCTT	SEQ ID NO:42
Exon D #1568	TTTCTCCTGGGACTCTGGGCAC	SEQ ID NO:43

**Table 2. STS/Dinucleotide Repeat Markers**

<b>Marker</b>	<b>GenBank Accession #</b>
D15S942	G04933
D15S1043	Z51622
D15S165	Z17271
D15S1031	Z51346
D15S1010	Z53401
D15S144	Z23286
D15S1007	Z53384
D15S995	Z53051
D15S1040	Z51533

Additionally, genomic P1 artificial chromosome (PAC) clones for  $\alpha 7$  were obtained from  
5 Genome Systems. PAC-64-A1 is 120 kbp long and contains both D15S1360 and the 5' end of  
the coding region. L76630 was localized in a genomic fragment containing the  $\alpha 7$  nicotinic  
receptor gene (CHRNA7), isolated from a human genomic library (Stratagene), by screening  
with a human  $\alpha 7$  cDNA clone (HP411).

A 6 kbp *EcoRI* genomic fragment was identified, partially sequenced, and found to include a CA dinucleotide repeat 3' of the last exon (GenBank Accession No. L76630). Flanking primers amplified 3 alleles (180, 178, 176 bp); allele frequencies were 0.06, 0.62, 0.32, with heterozygosity 0.51.

5        PCR was performed with 1.5 mM MgCl<sub>2</sub>: 94°C for 5 min, 20 cycles of 94°C for 1 min, 56°C for 2 min, 72°C for 1 min and 72°C for 5 min. The two polymorphisms were genetically mapped in 96 individuals from 6 reference families (Centre d'Etude du Polymorphisme Humain). These reference families were selected because they have three generations of individuals available for genotyping. Their DNA is available for genetic localization of markers,  
10       but their identities are confidential.

### EXAMPLE 3

#### Generation of Templates for Sequence Analysis of the Intron/Exon Borders.

15        In this Example, extra-long PCR (XLPCR), originally described by Barnes (Barnes, Proc Natl Acad Sci USA, 91:2216, 1994), was conducted using *rTth* polymerase with the Perkin Elmer XL/PCR kit (Perkin-Elmer), on a PTC 200 (MJ Research) thermal cycler with the following conditions: 94°C, 1 min, 1 cycle; 94°C, 15 sec/68°C, 10 min, 16 cycles; 94°C, 15 sec/68°C, 10 min 15 sec, 12 cycles; 72°C, 10 min, 1 cycle. Enzyme, primer concentration, and  
20       dNTP concentrations were as recommended by the manufacturer. A sublibrary of YAC b134H10 was constructed by *EcoRI* digestion and subcloning into Bluescript (SK-), (Stratagene), for splice junction determination on the larger introns.

To characterize the promoter and borders around exon 1 and 2, an *EcoRI* and *KpnI* sublibrary of PAC 25919 was constructed in Bluescript (SK-). A 2.9 kb clone containing exon 1,  
25       and a 5 kb clone containing exon 2 were identified by screening the PAC sublibrary by hybridization with an  $\alpha 7$  cDNA subclone containing 90 bp of 5' untranslated sequence, exon 1 and exon 2.

Tentative exon borders were deduced based upon the organization of the  $\alpha 7$  nAChR gene in the chick (Couturier *et al.*, Neuron, 5:847, 1990). Oligonucleotide primers, as shown in the  
30       table below, were designed from within the predicted exons that would amplify across the putative introns using extra-long PCR (XLPCR) with both genomic DNA and YAC b134h10

DNA. The exon primers used were as follows. For exon 5 to exon 10, the primers used were Primer #661 (TGACGCCACA TTCCACACTA A, SEQ ID NO:44); and Primer #591 (TTGTTTTCTT TCCACCAGTC A, SEQ ID NO:45). These primers amplify introns 5,6,7,8,and 9, with an approximate size of 14 kb. For exon 3 to exon 4, the primers used were  
5 Primer #1019 (CCAAGTTTTA ACCACCAACA TTTGG, SEQ ID NO:46); and Primer #1020 (TCCCCGCGGA AGAATGTCTG GTTCCAAAT CTG, SEQ ID NO:47). These primers amplify intron 3, with an approximate size of 8 kb.

The majority of intron-exon borders were determined from sequencing the XLPCR products. XLPCR products were not generated between exons 2 and 3 and between exons 4 and  
10 5, suggesting that these introns are large. Preliminary Southern blot data suggested that both are > 25 kb. The intron 2 acceptor border, and the intron 4 donor and acceptor borders were determined after sequencing *EcoRI* subclones derived from YAC b134h10. The intron 2 donor was determined from sequencing a *KpnI/EcoRI* fragment, subcloned from PAC 25919. Exon/intron border sequence and approximate lengths for introns and exons are summarized in  
15 Figure 1. All of the identified intron-exon borders are consistent with 5' donor and 3' acceptor RNA splice site consensus sequences.

The organization of the human  $\alpha 7$  nAChR gene was found to be identical to that found in chick with respect to number and size of exons. A signal peptide sequence predicted by  
20 13:596, 1993; Conti-Tronconi *et al.*, Proc Natl Acad Sci USA, 82:5208, 1985; and von Heijne, Nuc Acids Res, 14:4683, 1986), was found to be encoded by exon 1. Putative glycosylation sites (See *e.g.*, Séguéla *et al.*, *supra*, 1993; Schoepfer *et al.*, Neuron, 5:35, 1990) were found in exons 2, 4 and 5. The cysteine residues that form a putative disulfide bridge (Galzi *et al.*, Ann Rev Pharmacol, 31:37, 1991), were found to be encoded by exon 6. The vicinal cysteines at the  
25 acetylcholine (ACh) binding site, the  $\alpha$ -bungarotoxin binding site, and membrane spanning region I, are all coded by exon 7. Membrane spanning regions II and III (as in the rat) were found to be encoded by exons 8 and 9, respectively, while membrane-spanning region IV was found to be encoded by exon 10.

The putative promoter, and the borders for exons 1 and 2 were determined from  
30 sequencing *KpnI* and *EcoRI* subclones derived from PAC 25919, which contains exons 1-3 and sequences 5' of the coding region. A 2.9 kb *EcoRI-KpnI* fragment contained 2.6 kb of the region

5' of exon 1, exon 1 and 200 bp of intron 1. Sequence analysis indicated that 392 bp of the 5' region (GenBank Accession No. AF029837), shown in Figure 4, is 77% GC rich and lacks a consensus TATA box sequence. In this Figure, the nucleotides are numbered relative to the ATG translation initiation site (indicated with Met); the coding sequence is indicated in bold.

5 Consensus AP-2, Sp1, and CREB sequences are shown in boxes. Alignment of the chick (Matter-Sadzinski *et al.*, EMBO J, 11:4529, 1992) and human promoter sequences indicate they share only 52.9% homology. However, consensus Sp1, and AP-2 transcription factor binding sites are present in both human and chick  $\alpha 7$  promoters at approximately the same location, relative to the start of translation (Matter-Sadzinski *et al.*, *supra*, 1992). A CREB consensus  
10 binding sequence is present in the human promoter, but is not found in the chick.

The primers listed in Table 1 provide a means to obtain sequence information from genomic DNA. Using sequencing techniques standard in the art (*e.g.*, including, but not limited to standard dideoxy sequencing, chain termination sequencing using *Taq* DNA polymerase or other thermostable polymerases, and automated processes that use these and other technologies),  
15 the sequences near the intron and exon junctions can be obtained. Such primers have been successfully used to obtain sequence information from blood samples obtained from schizophrenic patients (*i.e.*, samples obtained as described in Example 1). Sequence obtained from this portion of the chromosome also finds use in providing linkage signal for other nicotine-dependent illnesses including, but not limited to, small cell lung cancer and juvenile myoclonic  
20 epilepsy. These sequences are then analyzed to determine if they contain pathogenic mutations that alter gene function by changing the amino acid coding, or by altering gene expression or response to promoter molecules, or by introducing variations in gene splicing. These mutant sequences are also expressed in transgenic cells in culture or in transgenic mice or in frog oocytes, to determine if they indeed cause altered gene function that produces heritable human  
25 illnesses such as schizophrenia.

#### EXAMPLE 4

##### Identification of Expressed Sequence Tagged cDNAs

In this Example, expressed sequence tagged (EST) cDNA clones were identified in the  
30 EST Database at the National Center for Biotechnology Information (NCBI), Bethesda, MD, by BLAST homology searches using  $\alpha 7$  cDNA specific sequences. Two clones (EST 3952 and EST

52861) were purchased from Research Genetics and sequenced bi-directionally as described in Example 5. Contigs were constructed using Sequencher software (Gene Codes).

## EXAMPLE 5

### Sequence Analyses and Restriction Mapping

In this Example, sequences were determined using standard sequencing kits and automated sequencing. In addition, genomic DNA probed with portions of  $\alpha 7$  cDNA was used to order *HindIII* restriction fragments.

### Manual Sequencing

PCR product for hand sequencing was prepared using the Exonuclease I-Shrimp Alkaline Phosphatase reagent pack (Amersham), per the manufacturer's directions. Sequencing was done using Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit from Amersham. The manufacturer's recommended component concentrations were used with 10 ng of template per 250 bp product per reaction. Reactions were run on a BioRad Sequi-Gene GT sequencing system (BioRad), using a 6% acrylamide/bisacrylamide (19:1) gel.

### Automated Sequencing

Plasmids to be sequenced were colony purified, using a Qiagen kit (Qiagen). PCR products from PACS, BACs, and YACs were gel purified using a Qiagen PCR product gel extraction protocol. Automated sequencing (ABI 373 or 377, Perkin Elmer) was conducted using Perkin Elmer ABI Dye Terminator or M13 Dye Primer kits, following manufacturer's protocols. Sequencing was organized into contigs using the Sequencher program (Gene Codes). All sequencing was bi-directional.

### Restriction Endonuclease Mapping

Southern analysis of genomic DNA probed with portions of  $\alpha 7$  cDNA was used to order *HindIII* restriction fragments (Dracopoli *et al.*, *supra*). DNA was transferred to Hybond N+, and hybridized at 40°C in 5X Denharts (0.5% SDS, 6X SSC and 50% formamide), then washed twice in 0.1% SDS and 0.1X SSC at 65°C for 10 minutes.

## EXAMPLE 6

### Large Insert Clone Contig

Total yeast DNA was isolated from YAC-bearing yeast using a spheroplast method (Dracopoli *et al.*, *supra*). Loci in and around the  $\alpha 7$  region were PCR amplified with loci specific primers (*i.e.*, primers shown in Table 2, as well as primers for D15S1360 described in Example 2). PCR was performed with 1.5 mM MgCl<sub>2</sub>: 94°C for 5 min, 20 cycles of 94°C for 1 min, 56°C for 2 min, 72°C for 1 min and 72°C for 5 min.

Mapping of specific exons was performed using the primers listed in Table 1 and the PCR conditions were 94°C for 2 minutes, 1 cycle; followed by 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and then 72°C for 30 seconds, followed by 72°C for 7 minutes-1 cycle.

Specific amplification was confirmed by sizing the products on agarose gel. PCR products from  $\alpha 7$  exons were excised from the gel, Qiagen extracted (Qiagen), and sequenced as described in Example 5.

Additional large insert genomic clones were isolated by PCR screening with  $\alpha 7$ -specific primers (Chumakov *et al.*, *supra*). YACs 953g6, 948a10, 853b12, and 969b11 were isolated from the CEPH YAC Library 3. PAC clones 64a1 and 25919 were identified by Genome Systems and BAC 467o18 was identified in a BAC library purchased from Research Genetics.

A tentative YAC contig was designed from markers in the YACs and information in the CEPH/Genathon Database. YACs providing linkage between the full-length and duplicated  $\alpha 7$  gene sequences, YACs 895f6, 776a12, 791e6, 811b6, 859c11, 801e1, 810f11, 966a4, 764f8, and 822g2, were obtained from Research Genetics. The contig, shown in Figure 5, was verified by PCR and sequencing of either  $\alpha 7$  sequence or published marker sequence. Loci from the 15q13-14 region were assigned to YACs, BACs, and PACs. The results confirmed the presence of markers previously assigned by Genethon (Human Genome Research Center; a publicly accessible database that maintains human genome linkage information). As indicated in Figure 5, two allele sizes for the L76630 loci were identified, suggesting that YAC 969B11 spans both  $\alpha 7$  nAChr loci.

Exons 5-10 of the  $\alpha 7$  nAChr gene and the polymorphic marker L76630 map to two distinct regions of the contig, suggesting a partial gene duplication. The distal, and full-length,  $\alpha 7$  nAChr gene maps close to D15S1360, as indicated by two PAC clones (64a1 and 25919) and one BAC clone (467o18). Both of these PACs, approximately 120 kb in size, contain the marker

D15S1360 which was used to demonstrate linkage of this region at 15q14 to a schizophrenic trait. Physical mapping of the  $\alpha 7$  gene <120 kb from the linkage marker suggested that the  $\alpha 7$  nicotinic receptor gene is an excellent candidate gene for this trait. The proximal duplicated exon sequences 5-10 of the  $\alpha 7$  nAChR gene map between D15S1043 and D15S165. The order of loci was determined to be D15S942, D15S1043, followed by the duplicated sequences L76630, exon 10, exon 9, exon 8, exon 7, exon 6, and exon 5, and then D15S165 and D15S1031. The closest marker flanking the 3' end of the  $\alpha 7$  nAChR gene could not be established and is either D15S1031 or D15S1010. Thus, the full-length gene with the 3'-end closest to D15S1031 has been tentatively oriented, based on the confirmed orientation of the duplicated sequences.

In order to determine if sequence differences were present that might distinguish duplicated exons 5-10 from the full-length gene, PCR products were generated and sequenced from 11 of the genomic YAC clones in the contig. Of these 11 clones, two (948a10 and 853b12) clearly mapped to the duplicated region between D15S1043 and D15S165, and eight mapped to the full-length  $\alpha 7$  nAChR gene region near D15S1360. All of the  $\alpha 7$  exons were found to be present in YACs 776a12, 791e6, 811b6, 953g6, b134h10, 859c11, 810f11 and 801e1. YAC 948a10 contained only exons 5-10, and 853b12, 6-10, while YAC 969b11 appeared to contain both loci. This YAC is 1.03 Mb in size (Figure 5), suggesting that the full-length  $\alpha 7$  gene and duplicated sequences are not more than 1 Mb apart.

Sequence variants found in DNA from duplicated and full-length genomic  $\alpha 7$  sequences are shown in Figure 2. In exon 6, a 2 bp deletion was identified at bases 497-498 (TG) in clones from the duplicated region, which results in a frame shift in the coding sequence and the insertion of a stop codon within the exon.

Additional sequence variants were found at bases 654, 793, 1269 and 1335 of the coding region. These are conservative base changes that do not change an amino acid. The polymorphic marker, L76630 is also duplicated as evidenced by the presence of a different number of CG repeats in the 3'UT of the full length  $\alpha 7$  gene and the 3' sequences following exon 10 in the duplicated sequences. YAC 969b11, which contains both full length and duplicated sequences also has two copies of L76630 as does a chromosome 15 hybrid, R379-2B2 (*See*, Figure 2).

## EXAMPLE 7

### RACE Analysis

In this Example, amino terminal clones for the human  $\alpha 7$  subunit were obtained by 5' RACE (*i.e.*, Rapid Amplification of cDNA Ends; *See*, Frohman, Amplifications 5:11, 1990),  
5 using a kit from Gibco-BRL, with some modifications. Although some of these products had the amino terminus nucleic acid sequences that were expected by homology with chicken and rat sequences, some had novel sequences that revealed the presence of unsuspected alternative exons. The present invention provides, for the first time, the sequences of these exons and their location in the genomic structure of  $\alpha 7$ .

10 Total RNA was isolated from normal human hippocampus by the method of Chomczynski and Sacchi (Chomczynski and Sacchi, Anal Biochem, 162:156, 1987). Briefly, brain tissue from the human hippocampus was disrupted in the presence of Solution D (4 M guanidium thiocyanate, 25 mM sodium citrate pH 7.0, 5 % sarcosyl, 0.1 M 2-mercaptoethanol) in a tissue homogenizer. The homogenized tissue was acidified with 0.1 X volume of 2 M  
15 sodium acetate, pH 4.0, with "X" referring to the initial volume of Solution D. The acidified tissue homogenate was extracted with 1X volume of water-saturated phenol and 0.2 volume of chloroform:isoamyl alcohol (49:1). The phases were separated by centrifugation (the supernatant contains RNA whereas the DNA and proteins remain in the interphase and the phenol). The RNA was precipitated by addition of an equal volume of isopropanol (20°C),  
20 followed by centrifugation. The RNA pellet was subsequently resuspended in 1 mM EDTA, pH 8.0. The concentration of the RNA was determined by measuring the absorbance at 260 and 280 nm.

The first strand cDNA synthesis for 5'-RACE was performed as indicated in the manufacturer's instructions, with the addition of methylmercuric hydroxide (7 mM) to reduce  
25 secondary structure. The cDNA was synthesized using a human gene-specific antisense oligonucleotide: 5'-AGGACCCAAA CTTCAG-3' (SEQ ID NO:48), complementary to 5'- sequence in the longest human clone from the primary cDNA screen. Following cDNA synthesis, terminal deoxynucleotide transferase was used to attach homopolymeric (dCTP) tails to the 3' ends of the cDNA. A nested gene specific antisense primer and an anchor primer from  
30 the 5'-RACE kit, both containing triplet repeat sequences for annealing to the pAMP1 vector, were used for PCR amplification of a homopolymeric, tailed cDNA product. The sequences of

the primers were: for the antisense primer, 5'-CAUCAUCAUC AUCCAGCGTA CATCGATGTA GCAGGAAGTC TTGAATAT-3' (SEQ ID NO:49), and the anchor primer 5'-CUACUACUAC UAGGCCACGC GTCGACTAGT ACGGGIIGGI IGGGIIG-3' (SEQ ID NO:50). In this anchor primer sequence, the "I" is inosine.

5 Briefly, the final composition of the PCR reaction for amplification of dC-tailed cDNA was as follows: 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 400 nM for both primers, 200 μM each dNTP, 8 % DMSO and 0.2 unit/μl *Taq* DNA polymerase. The PCR program was as follows: 94°C, 1 min; 57°C, 30 sec; 72°C, 2 min for 35 cycles; final extension at 72°C for 10 min, then soak at 4°C.

10 PCR products were Glassmax (Gibco-BRL) purified and reamplified with the same reaction conditions using the following program: 94°C, 1 min; 50°C, 30 sec; 72°C, 2 min for 5 cycles; 94°C, 1 min; 55°C, 30 sec; 72°C, 2 min for 35 cycles; extension at 72°C for 7 min, and soak at 4°C. Products from this PCR reaction were then gel purified and cloned into the pAMP1 vector (Gibco-BRL) with uracil DNA glycosylase according to manufacturer's directions, for  
15 subsequent automated sequencing, as described in Example 5.

A group of novel exons located in YAC, PAC and BAC clones containing the full-length gene and/or the duplicated α7 sequences was also evidenced. These novel exons were discovered in the process of comparing RACE clones, isolated during cloning of the α7 human cDNA, with EST cDNA clones (EST 3952 and EST 52861) found in the EST Database (NCBI)  
20 by homology screening. During cloning of the 5' end of the α7 coding region, the RACE technique was used to amplify the 5' end of the α7 cDNA (Frohman, *supra*, 1990). Although cDNA clones which matched sequence for published human α7 from a neuroblastoma cell line SH-SY5Y (Peng *et al.*, Mol. Pharm., 45:546, 1994), were obtained, clones with 5' sequence that could not be identified were also obtained.

25 When EST cDNA clones were subsequently found in the EST database by homology screening, several were identified that had exons 5-10 and unknown sequence 5' of exon 5. Comparison of the 5' ends of the RACE and EST products showed that the novel sequences are partially homologous. PCR primers were designed to these novel sequences for amplification from genomic DNA. Intronic sequence and consensus splice junctions that identified these  
30 sequences as four alternatively spliced and previously unreported exons were then identified. The sizes and splice junctions for these novel exons, designated as α7D, α7C, α7B, and α7A are

shown in Figure 6. In this Figure, the sequence of the RACE clone (GenBank Accession No. AF029838) is shown in uppercase, while intron boundaries are shown in lowercase, and are not included in the nucleotide numbering. The sizes of the exons are indicated below the exon designations. RACE clones, containing these novel exons were previously deposited with GenBank (RACE D-C-B-A-5-6; AF029838; RACE D-C-A-5-6, GenBank Accession No. AF029839).

## EXAMPLE 8

### RT-PCR Analysis

Total RNA was isolated from normal human hippocampus, human cingulate gyrus, the SH-SY5Y neuroblastoma cell line, and human immortalized lymphocytes with TRIzol reagent (Gibco-BRL) following manufacturer's instructions. The mutations seen in the PAC, BAC, YAC and published  $\alpha 7$  sequences were screened in seven normal subjects and SH-SY5Y cells. DNA was evaluated for all subjects, while cDNA was evaluated in exons 1-10 and exons 5-10 for all subjects, and exons D-10 were evaluated in one normal subject and SH-SY5Y cells. The DNA and RNA were obtained as detailed above. The cDNA was generated as previously detailed.

Total RNA was isolated from normal human hippocampus, human cingulate gyrus and SH-SY5Y neuroblastoma cell line by the TRIzol reagent (Gibco-BRL) following manufacturer's directions. RNA was stored as an ethanol precipitate until centrifugation and resuspended in 1 mM EDTA, pH 8.0 prior to cDNA preparation.

Total RNA was reverse transcribed at 42°C for 60 min in a 40  $\mu$ l volume with Superscript II reverse transcriptase (Gibco-BRL) and random hexamer primers (Pharmacia). The final concentration of the components of the reaction were as follows: 1x first strand buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM  $MgCl_2$ ), 1 mM dATP, 1 mM dGTP, 1 mM dCTP, 1 mM dTTP, 8  $\mu$ M random hexamers, 10 mM DTT, 0.5 U/ $\mu$ l placental Rnase inhibitor (Boehringer-Mannheim), 2.5 U/ $\mu$ l Superscript II reverse transcriptase and 500 ng of total RNA.

Primary PCR of the exon 1-10, exon 5-10 and exon D-10 products was performed using the Advantage-GC cDNA PCR kit (Clontech). Briefly, 5  $\mu$ l of RT products were diluted to 50  $\mu$ l with 40 mM Tricine-KOH, pH 9.2 at 25°C, 15 mM KOAc, 3.5mM  $Mg(OAc)_2$ , 5% DMSO, 3.75

µg/ml BSA, 0.2 mM of each dNTP, 0.2 µM of each primer, 1 M GC- Melt and 1x KlenTaq-1 DNA polymerase mix. Samples were incubated in a Perkin-Elmer 480 DNA Thermocycler.

For Exon 1-10, the sense primer was 5'-CGCTGCAGCT CCGGGACTCA ACATG-3' (SEQ ID NO:51), and the antisense primer was 5'-TGCCCATCTG TGAGTTTTC ACATG-3' (SEQ ID NO:52). The PCR conditions were 94°C, 1 min; 5 cycles at 94°C, 30 sec, 72°C, 3 min; 5 cycles at 94°C, 30 sec, 70°C, 3 min; 25 cycles at 94°C, 20 sec, 68°C, 3 min; final extension at 68°C, 3 min and soak at 4°C.

For Exon 5 to 3'UT α7 transcript, the sense primer was 5'-TGACGCCACA TTCCACACTA A-3' (SEQ ID NO:53), and the antisense primer was 5'-CCCCAAATCT CGCCAAGC-3' (SEQ ID NO:54). The PCR conditions were 5 cycles at 96°C, 1 min, 50°C, 30 sec, 72°C, 1 min; 30 cycles at 95°C, 30 sec, 62°C, 20 sec, 72°C, 30 sec; final extension at 68°C, 3 min and soak at 4°C.

For Exons D-10, the sense primer was 5'-CTCGGTGCCC CTTGCCATTT-3' (SEQ ID NO:55), and the antisense primer was 5'-CCTTGCCCAT CTGTGAGTTT TCCAC-3' (SEQ ID NO:56). The PCR conditions were 94°C 1 min, 5 cycles 94°C, 30 sec, 70°C, 3 min 5 cycles 94°C 30 sec, 68°C, 3 min, 25 cycles 94°C 20 sec, 66°C 3 min 1 cycle 68°C 3 min, cool to 4°C.

The products generated from exons 1-10, 5-10 and D-10 were further amplified to incorporate M13 primer sequences into products small enough to sequence in both directions. PCR conditions were as follows for all secondary, nested PCR amplifications. Perkin-Elmer Core reagents were used in standard concentrations using 2 mM MgCl<sub>2</sub>, 0.1 mM each dNTP, 1.5 U *Taq* Gold, 10% DMSO and 25 pM of each primer in a 50 µL reaction. PCR reactions were heated at 96° for 5 min, then 5 cycles were performed at 96°C for 1 min, 60°C for 30 sec, 72°C for 1 min; then 30 cycles for 95°C for 30 sec, 68°C for 20 sec, and 72°C for 30 sec, followed by a 7 min 72°C extension and cooling at 4°C.

All cDNA reactions were performed in duplicate using 50 ng RNA equivalents in a primary reaction, encompassing the full cDNA length of interest, then reamplified in nested, secondary PCR reactions to incorporate M13 primers into shorter products. DNA amplifications were performed in duplicate from 100 ng of needle-sheared template, within exon boundaries. The duplicates were then pooled, purified with a Centricon 100 (Amicon) column, and sequenced using standard M13 Dye Primer chemistry on an ABI 373 Automated sequencer. All templates were sequenced bi-directionally, except where sequence length did not allow a nested

primer. Alternate splice products were hand called from the electropherograms. Clean sequences were aligned and checked with Sequencher Software (Gene Codes Corporation).

DNA products were generated with primer pairs 1552/1553, 1101/1102, 1097/1098 and 1099/1100 to check the 497-498 deletion, 654/690, and 1269/1335 mutations, respectively.

5 These primers are shown in Table 3, below. In this Table, "1ry" and "2ry" refer to the first and second primer sets in nested PCR. The cDNA amplifications required three sets of primary amplifications, exons 1-10, exons 5-10 and exons D-10. Primer pair 1381/1382 was used to amplify exons 1-10; primers 1482/1483, 1101/1098 and 1099/1481 were then used as nested primers from this primary PCR to check 497-498, 654/693 and 1269/1335 respectively. The  
10 exon 5-10 product was amplified with primer pair 1502/1503, nested primers 1502/1483, 1101/1098 and 1099/1481 were used to check 497/498/654/693, 654/690 and 1269/1335 respectively. Exons D-10 were amplified with primers 1569/1562, and the nested primers 1553/1098 and 1097/1481 were used to check 497-498/654/690 and 1269/1335, respectively. Redundancy in the overlap of the secondary PCR products was used to double check some  
15 mutations, necessary when alternate splicing or base pair deletions occurred, making some base calls difficult.

Exon 3 codes for 15 amino acids near the amino terminal, in the extracellular domain. An alternate transcript without this exon appears in most PCR amplifications of this region, at a somewhat diminished concentration in comparison to the full-length transcript.

20 To determine if the exon 5-10 copy of  $\alpha 7$  was expressed, a second RT-PCR product was generated, encompassing only exons 5-10. The bases which appeared to be heterozygous in the DNA, but which are not heterozygous in the exon 1-10 transcript, are now fully accounted for in the 5-10 exon product, showing the exon 5-10 gene to be expressing as cDNA. The base changes fall into three categories, those seen only in the full-length 1-10 transcript, those changes  
25 present only in the 5-10 transcript and bases changes seen in both transcripts.

The TG deletion at 497-498 is only present in the 5-10 transcript; the C at 654 can be assigned to the 1-10 transcript, the T to the 5-10 transcript; the G at 933 can be assigned to both transcripts with an A in some subjects' 1-10 transcript and at 1335 the T can be assigned to the 1-10 transcript. The base changes seen at 690 and at 1269 appear to be present in both copies of  
30 the gene. These data are consistent with the base changes seen in the YAC, PAC and BAC clones, and the assignment of each clone to the duplicated or original gene.

**Table 3. Primer Sequences**

<b>Primer # Description</b>	<b>Sequence</b>	<b>SEQ ID NO:</b>
1097 sense/m13fwd+	CCCAGTACTTCGCCAGCACCATGAT	SEQ ID NO:57
1098 antisense/m13rev +	CCCCGTCGGGGTCGTGGTGGTGGTA	SEQ ID NO:58
1101 sense/m13fwd+	TCCCCGGCAAGAGGAGTGAAAGGTT	SEQ ID NO:59
1102 antisense/m13rev +	ACACCAGCAGGGCGAGGGCGGAGAT	SEQ ID NO:60
1099 sense/m13fwd+	GACCAGAGTCATCCTTCTGAACTGG	SEQ ID NO:61
1100 antisense/m13rev +	TTTCAGGTAGACCTTCATGCAGACA	SEQ ID NO:62
1553 sense/m13fwd+	CGATGTACGCTGGTTTCCCTTTGAT	SEQ ID NO:63
1552 antisense/m13rev +	TTCCCACTAGGTCCCATTCTCCATT	SEQ ID NO:64
1382 sense/1ry cDNA	CGCTGCAGCTCCGGGACTCAACATG	SEQ ID NO:65
1381 antisense	TGCCCATCTGTGAGTTTCCACATG	SEQ ID NO:66
1502 sense/1ry cDNA	TGACGCCACATTCCACACTAA	SEQ ID NO:67
1503 antisense	CCCCAAATCTCGCCAAGC	SEQ ID NO:68
1569 sense/1ry cDNA	CTCGGTGCCCCTTGCCATTT	SEQ ID NO:69
1562 antisense	CCTTGCCCATCTGTGAGTTTCCAC	SEQ ID NO:70
m13 sense/extension	TGTAAAACGACGGCCAGT	SEQ ID NO:71
m13 antisense/extension	CAGGAAACAGCTATGACC	SEQ ID NO:72
1482 sense/m13fwd+/2ry	AAGGAGCTGGTCAAGAACTACAATC	SEQ ID NO:73
1483 antisense/m13rev+	CCGGAATCTGCAGGAAGCAGGAACA	SEQ ID NO:74
1101 sense/m13fwd+/2ry	TCCCCGGCAAGAGGAGTGAAAGGTT	SEQ ID NO:59
1098 antisense/m13rev +	CCCCGTCGGGGTCGTGGTGGTGGTA	SEQ ID NO:58
1502 sense/2ry cDNA	TGACGCCACATTCCACACTAA	SEQ ID NO:67
1483 antisense/m13rev+	CCGGAATCTGCAGGAAGCAGGAACA	SEQ ID NO:74
1553 sense/m13fwd+/2ry	CGATGTACGCTGGTTTCCCTTTGAT	SEQ ID NO:63
1098 antisense/m13rev +	CCCCGTCGGGGTCGTGGTGGTGGTA	SEQ ID NO:58
1097 sense/m13fwd+/2ry	CCCAGTACTTCGCCAGCACCATGAT	SEQ ID NO:57
1481 antisense/m13rev+	CCAGGCGTGGTTACGCAAAGTCTTTG	SEQ ID NO:75
1099 sense m13fwd+/2ry	GACCAGAGTCATCCTTCTGAACTGG	SEQ ID NO:61
1481 antisense/m13rev+	CCAGGCGTGGTTACGCAAAGTCTTTG	SEQ ID NO:75

An RT-PCR product was generated from exon  $\alpha 7D$  to exon 10 from one normal brain and from SH-SY5Y cells. The resulting cDNA product contained alternate splice products with exons shown in Figure 2. The 2 base pair deletion seen at bases 497-498 in the DNA that is not present in the exon 1-10 transcript was seen in the D-10 transcript, while all of the D-10 product in SH-SY5Y was deleted at 497-498, and subject SL061 was heterozygous for the deletion in the D-10 product. The presence of the T at base 757 connects this base change to the TG deletion. The G at 690 was not expressed in either cDNA. The A at 933 was not present in the minus TG strand of SH-SY5Y. The T at 1296 was expressed in subject SL061. These products, in subject SL061 cannot differentiate between the exon 5-10 product splicing to exon D versus exon 1, however the product in D-10 from SH-SY5Y can, since only the minus TG strand was expressed, negating the possibility that exons 5-10 from the 1-10 gene are splicing to exon D.

These new exons have been designated as 3' $\alpha 7A$ ,  $\alpha 7B$ ,  $\alpha 7C$ ,  $\alpha 7D$  5'. The RACE products were variable in their inclusion of Exon B, similar to the EST clones. However, PCR products including exons D-10 gave many alternate splice products between exons D, C, B, 5 and 6. This same phenomenon was seen in the exon 1-10 transcripts between exons 2 and 6. Based on these results, it was not possible to fully evaluate whether any of the D-10 transcript contain only exons 5-10 from the duplicated region or if this transcript contains some splicing of 5-10 from the 1-10 full gene sequence, since the cell line and the brain gave differing results. Subcloning is used to fully evaluate the base changes to separate the various splice products.

These results indicate that the primer sequences described herein can be successfully used to screen both genomic DNA and mRNA for the presence in DNA and the expression in mRNA of sequences which are polymorphic (*i.e.*, different) between individuals. Standard automated and manual sequencing methodologies are used to locate differences in samples obtained from individuals. It is contemplated that some of these polymorphisms, as well as others, have pathogenic roles. These polymorphisms are also used to relate the inheritance of specific alleles of  $\alpha 7$  genes through families to the presence of illness or physiological dysfunction, using standard methods known in the art for linkage analysis.

## EXAMPLE 9

### Single Strand Conformation Polymorphism (SSCP) Analysis

PCR products, <200 bp, containing a single sequence variant were amplified with <sup>33</sup>Pγ-

ATP kinased primer sets using Promega T4 kinase as known in the art (*See e.g., Dracopoli et al., supra*). The primers used in this Example were:

**Table 4. Primers Used for SSCP Analysis**

Exon and Primer #	Sequence	SEQ ID NO:
Exon 6b / #1243	GATGTGCAGCACTGCAAACAA	SEQ ID NO:76
Exon 6b / #1144	TTAAAGCTTGCCCAGGAATAGG	SEQ ID NO:77
Exon 6d / #1124	GGAAGTGTGTGTATTTTCAGC	SEQ ID NO:78
Exon 6d / #1245	AAGACCAGGACCCAAACTTGT	SEQ ID NO:79
Exon 7d / #1143:	GCTTGTGTGTGGTATACACATTG	SEQ ID NO:80
Exon 7 / #675	GTAGAGTGTCTCTGCGGC	SEQ ID NO:81
Exon 10 (1438)/ #672	GGTCCGCTACATTGCCAA	SEQ ID NO:82
Exon 10 / #593	TGATGGTGAAGACCGAGAAGG	SEQ ID NO:83

Products, denatured with loading dye (7.26 M urea, 60% formamide, 22 mM EDTA, 32 mM NaOH, 0.25% bromophenol blue, 0.25% xylene cynol), were analyzed on GeneAmp detection gels (Perkin Elmer) run at both 6°C and 25°C, using Bio Rad PowerPac 3000 with a temperature probe, as described by the manufacturer.

Thus, the frequency of these sequence variants was examined, using SSCP in a group of 43 normal control subjects with no history of mental illness. Primer sets derived from the exon and intron-exon boundary sequences are used to amplify 200 bp portions of the gene from individuals with schizophrenia and their relatives, in order to identify sequence changes that affect gene function. Sequence changes that are not known to affect gene function, but can serve as markers to trace heritability of particular gene regions through families, are also identified in this process. The -2 bp deletion and the heterozygosities at 654, 690, 1269, at 1335 were found in this Example.

Almost all subjects were heterozygotic at positions 654 and 690. Nucleotide positions 1269 and 1335 were also found to be polymorphic, suggesting that the duplicated sequences have diverged since the duplication event.

These results indicate that the primer sequences described herein can be successfully used to screen genomic DNA in SSCP, a standard genome screening technique, for polymorphic differences in DNA sequences between individuals. It is contemplated that these polymorphisms, as well as others, have pathogenic roles. These polymorphisms are also used to relate the inheritance of specific alleles of  $\alpha 7$  genes through families to the presence of illness or physiological dysfunction, using standard methods known in the art for linkage analysis.

## EXAMPLE 10

### Electrophysiological Recording, Linkage Analysis, and Nonparametric Methods

Electroencephalographic activity was recorded at the vertex and electrooculographic activity was recorded from the superior orbital-lateral canthus. Five averages of sixteen responses each to paired clicks were obtained, using standard methods (*See, Griffith et al., Psychophysiology*, 32:460, 1995). The P50 responses were distinguished from pre-stimulus activity for both normals and schizophrenics at a high level of significance ( $P < 0.001$ ). The averages were reviewed by two investigators, blind to genetic information, who rejected any average containing excessive electrooculographic activity, drowsiness, startle, or other artifacts; the remainder were combined into a grand average, from which the P50 amplitudes were measured and their ratio (second response/ first response) was calculated automatically by a computer algorithm (Nagamoto *et al.*, *Biol Psychiat*, 25:549, 1989). Seven subjects were not used, because artifact-free averages could not be selected from their recordings. Recordings were initially performed, then repeated approximately three years later. The earlier recordings were reanalyzed for 2 subjects who were later deceased, for 10 subjects who refused repeat recording, and for 2 patients who were later on atypical neuroleptics, which can normalize the P50 ratio; other neuroleptic medication do not affect the phenotype (Nagamoto *et al.*, *Biol Psychiat*, 40:181, 1996).

Parameters for lod score analyses of P50 ratios were determined from the distribution of values in 43 unrelated normal individuals and 36 unrelated schizophrenic patients (Waldo *et al.*, *Schizophr Res*, 12:93, 1991) and from observations of the segregation of P50 ratios in the nine

multiplex schizophrenic families (*i.e.*, the families described in Example 1). Elevated P50 ratios were defined as values greater than or equal to 0.50, which were found in 91% of the unrelated schizophrenics and 6% of the normals. Of the remaining unrelated schizophrenics, most had values between 0.41 and 0.49, a range therefore coded unknown for the linkage analysis. If this  
5 unknown range was extended to include values between 0.40 and 0.60, the results were changed substantially (*e.g.*, lod scores were decreased by an average of 0.54 across the markers in the 15q13-14 region due to the loss in information). For lod score analyses, frequency of a gene for abnormal P50 ratio was fixed at 0.05, penetrance for the normal genotype was fixed at 0.01, and penetrance for the abnormal genotypes was fixed at 0.8 (Coon *et al.*, Biol. Psychiat., 34:277,  
10 1993). These parameters result in a morbidity for abnormal P50 ratio of 8.7% and a phenocopy rate among abnormal subjects of 10.4%. The FASTLINK version of the LINKAGE program was used to compute lod scores at various recombination fractions,  $\Theta$  (Lathrop *et al.*, Proc Natl Acad Sci USA, 81:3443, 1984). No significant heterogeneity was found using the HOMOG program (Ott, *Analysis of Human Genetic Linkage*, Johns Hopkins Univ. Press, Baltimore,  
15 1991). The chance of false positive lod score results was determined using SLINK (Ott, Proc Natl Acad Sci USA, 86:4175, 1989); 1000 replicates of the pedigrees were simulated, assuming no linkage to the marker under analysis. Lod score analysis was performed for each replicate under the dominant model; the highest score observed for D15S1360 and P50 under the assumption of no linkage was 1.87.

20 Sibling pair analysis was performed using the SIBPAL program (Elston, *SIBPAL, Statistical Analysis for Genetic Epidemiology*, Louisiana State Univ. Medical Center, New Orleans, LA, version 2.2, 1995). Marker data were used to estimate the proportion of alleles shared through a common ancestor (*i.e.*, identical by descent) for each possible sibling pairing within the linkage families. A test was performed to determine if the proportion of alleles shared  
25 was > 0.50 for abnormal/abnormal pairs. To calculate *P* values, 1000 replicates of the 9 families were simulated for each marker to determine empirical distributions. Degrees of freedom were adjusted downward for non-independence when multiple pairings were used from the same sibship within a family.

A newly developed method, Nonparametric Linkage, uses information from all  
30 genotyped members of a pedigree to assess the extent of alleles shared identical by descent among all affected individuals. The resulting statistic was normalized, first by subtracting the

expected sharing score under the null hypothesis (no linkage from the observed score), and then dividing by the score variance under the null hypothesis. Thus the statistic is asymptotically distributed as a standard normal variable (Z score) under the null hypothesis. Calculations of Nonparametric Linkage statistics were carried out using the GENEHUNTER computer programs (Elston, *supra*). GENEHUNTER also uses an improvement to a previously described algorithm to perform complete multipoint linkage analysis with a large number of highly polymorphic markers in pedigrees of moderate size (Kruglyak *et al.*, Am J Hum Genet, 58:1347, 1996). Due to computational constraints, the three largest pedigrees were each split into two parts.

Only one marker, D15S1360, yielded a lod score  $>3.0$  (lod score maximum = 5.3, theta = 0.0,  $P < 0.001$ ). DNA markers flanking D15S1360 also gave positive lod scores. Multipoint analysis showed a maximum lod score at D15S1360 of 5.29. Both the sibpair analysis and nonparametric linkage analysis gave confirming positive results of similar statistical significance. The sibpair analysis showed 0.70 proportion of D15S1360 alleles among siblings with abnormal P50 ratios ( $T = 4.07$ ,  $P < 0.0005$ ). Two point results from the non-parametric analysis were most significant for D15S1360 ( $Z = 3.95$ ,  $P < 0.0002$ ). A complete multipoint analysis using nine chromosome 15q markers gave a maximum value at D15S1360 ( $Z = 5.04$ ,  $P < 0.000016$ ).

From the above it should be clear that the present invention provides gene sequences encoding mammalian  $\alpha 7$  genes and proteins. The present invention further provides compositions and methods for targeted therapy directed to  $\alpha 7$  abnormalities.

## EXAMPLE 11

### Refinement of the Physical Map of the P50-Schizophrenia Linkage Region

This example provides details of further physical mapping of the region of chromosome 15q13-q14 that is inherited in subjects with the P50 deficit and with schizophrenia. The contig depicted in Figure 11 includes multiple bacterial artificial chromosomes, and map locations for additional expressed sequence tags and markers. The region is defined by 30 markers and is estimated to be about 4 Mb in length. The full-length *CHRNA7* gene, implicated in the P50 deficit in schizophrenia, is localized at this site between unique markers D15S1013 and D15S1010. Mapping of  $\alpha 7$  exons showed that exons 5 to 10 of *CHRNA7* had been duplicated, along with a large cassette of DNA containing several other genes, and inserted proximal to the

full-length gene. The insertion occurred next to novel exons D-C-B-A, with which the duplicated  $\alpha 7$  exons are expressed as messenger RNA (*dupCHRNA7*; GenBank Accession No. AF029838). The novel *dupCHRNA7* transcript was detected in multiple tissues, including human brain and blood leukocytes.

During development of the present invention, exons D-C-B-A were found to be both duplicated and expressed with downstream sequences that are not of  $\alpha 7$  origin (GenBank Accession No. AA861176). These novel exons were also found to map on chromosome 3 by hybrid clone panel analysis. It is contemplated that exons D-C-B-A, contained in clone AA861176, were duplicated at least once on chromosome 15, with one insertion site near the dinucleotide repeat (D15S1043), before the partial duplication of the *CHRNA7* gene. Ultimately, the large cassette containing  $\alpha 7$  exons 5 to 10, dinucleotide repeat L76630, and expressed sequence tag WI13983 was duplicated and inserted proximally, interrupting the duplication of AA861176. Additional analysis of exons D-C-B-A indicated that exon D actually contains 2 exons and an intervening sequence. The two newly defined exons are designated as D' (proximal) and D (distal). The unique DNA sequence between the full-length *CHRNA7* gene and *dupCHRNA7* is approximately 1 Mb, and contains a large number of mapped expressed sequence tags and markers. The site of the marker D15S1360, isolated from a YAC containing *CHRNA7*, has been more precisely mapped to intron 2, by examining the sequence available from the National Human Genome Research Institute (bacterial artificial chromosomes 717i24 and 198g2). The D15S1360 repeat has been used extensively for genotyping of both schizophrenic patients and controls in the studies disclosed herein. Only two alleles were ever observed in any one individual. Furthermore, the promoter and exons 1 to 4 of the full-length  $\alpha 7$  gene were found only in bacterial artificial chromosomes and P1 artificial chromosomes containing D15S1360, all of which map between D15S1031 and D15S1040. Thus, during development of the present invention, the region 5' of exon 4, containing the promoter region of the full-length  $\alpha 7$  gene, was determined not to be duplicated.

## EXAMPLE 12

### Subject Selection and Sample Collection for *CHRNA7* Promoter Analysis

Subjects were analyzed in a modified case-control study for polymorphisms in the core promoter of the full length *CHRNA7* gene. A total of 298 schizophrenic subjects were available

for screening (See, Table 5). The sample contained 188 subjects from the NIMH Schizophrenia Genetics Initiative, including DNA samples from 20 families used in a sib-pair analysis positive for schizophrenia (Leonard *et al.*, Am J Med Genet, 81:308-312, 1998). These DNA samples were derived from lymphoblast cultures in the NIMH collection. Three schizophrenic lymphoblast cultures were obtained from Israel, while the remaining DNA samples were isolated from either postmortem brain or lymphoblasts collected in the Denver Schizophrenic Center (Denver, CO). The samples collected in Denver included 25 specimens from patients with childhood onset schizophrenia.

**Table 5. Subjects Used For Screening of the *CHRNA7* Promoter\***

Source	Subjects	Number	DNA Source	P50 Ratio
NIMH	SZ	188	lymphoblasts	0
Denver	SZ	49	lymphoblasts	34
Denver	SZ	33	brain	0
Denver	COSZ	25	lymphoblasts	18
Israel	SZ	3	lymphoblasts	0
<b>Total</b>	<b>SZ</b>	<b>298</b>	<b>both</b>	<b>52</b>
Denver	Control	152	lymphoblasts	151
Denver	Control	13	brain	0
<b>Total</b>	<b>Control</b>	<b>165</b>	<b>both</b>	<b>151</b>

\*Abbreviations: NIMH, National Institute of Mental Health Genetics Initiative for Schizophrenia; SZ, schizophrenia; and COSZ, childhood-onset schizophrenia.

Postmortem brain was donated by the family of the deceased through the Colorado Uniform Anatomical Gift Act (1968) and collected at autopsy. Hospital and autopsy records were reviewed, and family members and physicians were interviewed to determine age, sex, cause of death, and mental illness status. Brains were weighed, examined for gross pathological features, and divided sagittally. One hemisphere was preserved in formalin for neuropathological analysis. The other hemisphere was sliced coronally into 1-cm slices, from which multiple regions were dissected in blocks, frozen in dry-ice snow, and packaged for

storage at  $-80^{\circ}\text{C}$  (Leonard *et al.*, Biol Psychiatry, 33:456-466, 1993). DNA was isolated from cortex, by means of standard methods (Sambrook *et al.*, eds., Molecular Cloning: A Laboratory Manual, 3<sup>rd</sup> edition, Cold Spring Harbor, NY: Cold Spring Harbor Press, 2001). Of the 165 control DNA samples used in the study, 152 were isolated from blood collected in the Denver Schizophrenia Center and had no evidence of current or past psychosis as determined with a Structured Clinical Interview for Axis I DSM-IV Disorders-Non-Patient Edition (SCID-I/NP), Version 2.0 (First *et al.*, Biometrics Research Department, New York State Psychiatric Institute, 1996). In addition, these controls had a Family History Research Diagnostic Criteria, 3<sup>rd</sup> edition, interview that showed no evidence of family history of psychosis (Endicott *et al.*, Research Assessment and Training Unit, New York State Psychiatric Institute, 1978). All local subjects included in this study provided written informed consent by means of forms approved by the University of Colorado Health Sciences Center Internal Review Board.

Auditory evoked potentials were recorded on 151 of the living controls, by published methods (Freedman *et al.*, Schizophr Res, 4:233-243, 1991). Briefly, auditory sensory gating is measured by means of the P50 wave of the electroencephalogram response to paired auditory stimuli delivered in the form of clicks. After the second stimulus, delivered 0.5 second after the first, the P50 response is decreased in normal individuals. In most schizophrenic subjects, the response to the second stimulus is not as greatly diminished as in controls; in some subjects the second response is larger than the conditioning response. Control subjects with no history of mental illness, generally have P50 ratios of the T/C response amplitudes that are less than 0.50. Although some P50 ratios in controls are higher, before the development of the present invention, it was not known what causes this variation.

Ethnicities of all subjects were recorded from self-report or family interview and represented three major groups. White subjects accounted for approximately 65% of the samples from schizophrenic patients and 61% of the controls, and African Americans approximately 31% of the schizophrenic sample and 34% of the control subjects. Hispanics accounted for 4% of samples from schizophrenic patients and 5% of the controls. All schizophrenic subjects in each family were screened for polymorphisms to detect the possible presence of different variants in related individuals.

## EXAMPLE 13

### Mutation Screening of the *CHRNA7* Promoter

The  $\alpha 7$  gene cluster in the 15q13-q14 linkage region was selected as the most likely candidate gene group for mutation studies, based on inclusion of the linkage marker D15S1360 within intron 2 of the full-length nAChR $\alpha 7$  gene and the neurobiological evidence described herein that is consistent with diminished  $\alpha 7$  expression or function. Because nonsynonymous changes in the coding region were found to be rare and not associated with schizophrenia, the promoter region of the gene was first examined.

Genomic DNA was isolated from individuals as previously described (Gault *et al.*, Genomics, 52:173-185, 1998), and 231 bases proximal to the  $\alpha 7$  ATG translation start site were screened. Single-stranded conformational polymorphism (SSCP) analysis and sequence analysis were used to identify polymorphisms in this core promoter region (Gault *et al.*, *supra*, 1998). Briefly, two primers sets for overlapping fragments covered the region from bases -14 to -268 (primer sets 4 and 5 of Table 6). The primer sets were phosphorylated with [ $\gamma$ -<sup>33</sup>P]-adenosine triphosphate and T4 kinase (Promega), then used separately to amplify the promoter region by PCR. PCR was done with *Taq* Gold and a GeneAmp PCR System 9600 kit (Perkin-Elmer) using the following program: 95°C, 3 min; 95°C, 30 sec; 60°C, 30 sec; 72°C, 30 sec; for 35 cycles; then 72°C, 3 min. The products were denatured with loading dye (7.26M urea, 60% formamide, 22mM EDTA, 32mM sodium hydroxide, 0.25% bromophenol blue, and 0.25% xylene cyanol) and were separated on GeneAmp detection gels (Perkin-Elmer) run at 4°C and 25°C by means of a Bio-Rad Power Pac 3000 Power Supply with a temperature probe. The results were similar at both temperatures. SSCP analysis of DNA samples, from both schizophrenic and control subjects, was completed in the same experiment.

**Table 6. Primer Sets Used For PCR, DNA Sequencing and SSCP Analysis\***

Sets	Sequence	Size (bp)	Base	Use	SEQ ID NO
1 S	5' -GGTTGGCAAGACTTCCGAAGCC-3'	618	-553 to -531	PCR, SEQ	126
1 AS	5' -GTGGCTTTACCGTGCAGGAGCG-3'		+44 to +65	PCR, SEQ	127
2 S	5' -AGTACCTCCCGCTCACACCTCG-3'	271	-269 to -248	PCR, SEQ	128
2 AS	5' -ATGTTGAGTCCCGGAGCTGCAG-3'		-20 to +2	PCR, SEQ	129
3 S	5' -CTGGCCAGAGGCGCGAGGCCG-3'	N/A	-347 to -327	SEQ	130
4 S	5' -GGGGCTCGTCACGTGGAGAGGC-3'	180	-170 to -149	SSCP	131
4 AS	5' -AGCAGCGCATGTTGAGTCCCGGAGC-3'		-14 to +10	SSCP	132
5 S	5' -GTACCTCCCGCTCACACCTC-3'	176	-268 to -249	SSCP	133
5 AS	5' -CGGCTCGCGCGCCTTTAAGGA-3'		-112 to -92	SSCP	134
6 S	5' -AGTACCTCCCGCTCACACCTCG-3'	696	-269 to -248	PCR, SEQ	135
6 AS	5' -GGAGGCTCAGGGAGAAGTAG-3'		+407 to +427	PCR, SEQ	136

\*Abbreviations: S, sense; AS, antisense; and bp, base pairs.

Automated DNA sequencing on an Applied Biosystems 377 DNA Sequencer was used for verification of polymorphisms and determination of the specific bases changes, as previously described (Gault *et al.*, *supra*, 1998). Generally, a large fragment of 618 bp was generated with the use of primer set 1. A final concentration of 1.25M betaine (Sigma-Aldrich), added to Master Mix 2 in the Expand Long Template PCR System kit (Roche) was used to amplify the fragment, with the following PCR program: 93°C, 2 min; 38 cycles at 93°C, 30 sec; 62°C, 30 sec; 72°C, 1 min; followed by 72°C, 7 min. Briefly, 200 ng of genomic DNA was diluted in a volume of 25 µl to the following final concentrations: 1X Expand Long Template PCR Buffer 3 (Roche), containing 0.75mM magnesium chloride, 1.67 U of Expand Long Template enzyme mixture (*Taq* and *Pwo* thermostable DNA polymerases), 0.25mM of each deoxynucleotide triphosphate, 0.4µM of each primer, and 1.25M betaine. An additional primer set 2 was often used for sequencing of a shorter fragment in the proximal promoter region (271 bp). The PCR conditions for the shorter fragment were the same as for the longer fragment.

The G/C(-194) and G/A(-191) variants had indistinguishable SSCP patterns. Samples with these polymorphisms were analyzed with WAVE technology (Transgenomics). WAVE detects sequence changes in PCR products based on differential separation through temperature-modulated liquid chromatography and a high-resolution matrix with detection by absorbance at 254 nm. During development of the present invention, the PCR products generated with primer set 2 were used. An aliquot of the PCR fragment generated from control or patient DNA was then used for heteroduplex formation in the thermal cycler as follows: 95°C for 5 min, ramp slowly from 95°C down to 25°C for 45 min, then hold at 4°C. The melting profile of a normal 271 bp promoter sequence was determined with the Wavemaker Program (Transgenomics). A temperature curve was generated for the heterozygous samples containing either G/C(-194) or G/A(-191) at temperature ranging from 69°C to 73°C. The resulting chromatograms showed the presence of heteroduplex peaks that were resolved optimally at 71°C. A triethylammonium acetate and acetonitrile gradient specified by the manufacturer was used for elution. All subsequent samples were run under identical conditions.

Approximately 2.6 kb (SEQ ID NO:122) of DNA sequence 5' of exon 1 in the full-length *CHRNA7* gene was cloned from P1 artificial chromosome 24919 that contains *CHRNA7* exons 1 to 3 (Incyte). Subclones of this region were constructed for determination of functional domains for gene transcription (See, Figure 12, panel A). Base pair numbering begins with -1 at the position preceding the translation start in exon 1. The three fragments indicated were cloned into the pGL3 Basic Vector (Promega) for analysis of promoter sequence effects on the reporter gene luciferase. A fragment of 231 bp, immediately 5' of exon 1, was identified as the core promoter sequence and is sufficient to drive high levels of transcription in vitro. Sequences further upstream, included in fragments of 1.0 kb and 2.6 kb, were identified as containing putative repressor elements.

The 231-bp core promoter region is homologous to the bovine  $\alpha 7$  core promoter region, including conservation of some transcription factor consensus sequences (Carrasco-Serrano *et al.*, J Biol Chem, 273:20021-20028, 1998). Thus, the human  $\alpha 7$  promoter region is contemplated to be regulated in part by Sp1 and AP-4 transcription factors, for which there are 2 clusters of consensus sites (See, Figure 2, panel B). The regions including the Sp1 sites were also identified as G/C boxes, which are contemplated to bind other transcription factors. There is a

consensus serum responsive element (SRE), also found in the bovine gene, but not in chick gene (Couturier *et al.*, Neuron, 5:847-856, 1990).

Mutation screening was completed for the 231-bp core promoter in 195 schizophrenic individuals and 165 controls, demonstrating a complex cluster of variants (*See*, Table 7). There were 12 different single nucleotide changes, including two insertions and a deletion. Many of the variants lie in putative transcription factor consensus binding sequences (*See*, Figure 2, panel B). For instance, the G/C variant at -194 introduces a new SpI site. In addition, some subjects were found to carry double variants that were combinations of the single variants (8 different combinations). The total numbers of single and double variants found in control and schizophrenic subjects are shown in Table 7 and Table 8, respectively, stratified by ethnicity. One polymorphism, an insertion of +CGGG at -140 bp, was found in a single subject with a diagnosis of psychosis, not otherwise specified (DSM-IV, 298.9). As this diagnosis was not included in either the control or schizophrenic sample diagnoses, this individual was not included in Table 7 or in the statistical analysis, but is disclosed to indicate that additional and perhaps more complex polymorphic patterns may remain to be discovered with the methods and compositions disclosed herein. Forty-seven of 165 control individuals and 71 of 195 schizophrenic patients had one of the single polymorphisms. Although one single variant (-93 bp) and two double variants (-93 bp/-194 bp and -191 bp/-194 bp) were found only in control subjects, a larger number of both single and double variants were found in schizophrenic patients than in controls. The difference was not, however, statistically significant. Eight of the 12 variants (-86 bp, -92 bp, -143 bp, -178 bp, 480 bp, -191 bp, -194 bp, and -241 bp), marked with asterisks in Table 7, were found to be more prevalent in schizophrenic subjects. Twenty-seven of 165 control subjects had one of these 8 variants, but 59 were found in the 195 schizophrenic patients. Association of the single variant -86 bp C/T with schizophrenia in the combined ethnic groups reached significance ( $P=0.04$ ). This polymorphism was examined alone because -86 bp C/T was found to be the most common variant in the region, and because it was found to have the highest prevalence in schizophrenic patients. It is found more frequently in whites than in African Americans. The genotype relative risk for this variant was 2.39 (95% confidence interval, 1.07-5.32). The principal polymorphisms found in African American schizophrenic patients were the G deletion at -178 bp and the G/A substitution at -191 bp. Although more

variants at these sites were found in schizophrenic subjects than in controls, fewer subjects were carrying each of the polymorphisms and the differences were not significant.

Thirty-four affected full sib pairs were examined in 30 families that had one or more of the promoter polymorphisms. Fourteen of the 34, or 0.41 sib pairs, shared at least one of these variants. For the common -86bp C/T variant, 6 of 12 sib pairs shared the polymorphism.

**Table 7. Single Promoter Variants in Control and Schizophrenic Subjects**

Variant	Control Subjects				Schizophrenic Subjects			
	White	African	Hispanic	Total	White	African	Hispanic	Total
-46 G/T	0	11	0	11	0	9	0	9
-86 C/T	9	0	0	9	20	1	2	23*†
-92 G/A	1	0	0	1	1	1	0	2*
-93 C/G	1	0	0	1	0	0	0	0
-143 G/A	0	0	0	0	0	1	0	1*
-172 +CGGGGG	1	0	0	1	0	0	0	0
-178 -G	0	3	0	3	0	9	0	9*
-180 G/C	0	0	0	0	0	0	1	1*
-190 +G	0	5	2	7	1	2	0	3
-191 G/A	0	3	0	3	1	6	0	7*
-194 G/C	9	2	0	11	12	1	1	14*
-241 A/G	0	0	0	0	2	0	0	2*
Total Variants	21	24	12	47	37	30	4	71
Total Subjects	103	54	8	165	129	56	10	195

\*Found more frequently in schizophrenic subjects. †P=0.04.

**Table 8. Double Variants in Control and Schizophrenic Subjects**

	Control Subjects				Schizophrenic Subjects			
Combination	White	African	Hispanic	Total	White	African	Hispanic	Total
-46/-178	0	0	0	0	0	3	0	3*
-46/-190	0	0	0	0	0	1	0	1*
-46/-191	0	0	0	0	0	1	0	1*
-86/-194	2	0	0	2	1	0	0	1
-86/-241	0	0	0	0	1	0	0	1*
-93/-194	1	0	0	1	0	0	0	0
-178/-191	0	0	0	0	0	1	0	1*
-191/-194	0	1	0	1	0	0	0	0
Total Variant	3	1	0	4	2	6	0	8†
Total Subjects	103	54	8	165	129	56	10	195

\*Found only in schizophrenic individuals. †P=0.38 (not significant).

In other embodiments, mutation screening of the 2.6 kb upstream regulatory region is done with a Transgenomics WAVE™ denaturing high performance liquid chromatography system (DFPLC). This system detects pattern differences in PCR fragments bearing mutations. Primers are designed for overlapping fragments of approximately 300-500 bp from -2600 bp to the proximal promoter. The size of the fragment ranges from 100 bp to 300 bp and depends upon the melting profile for the sequence, as determined by utilization of the Transgenomics software. At least 10 fragments are screened. The fragments generated from each subject are then run on the DFPLC system. Fragments showing a pattern different from the wild-type are sequenced for identification of the specific mutation. Patterns similar to wild type are mixed with a wild type sample to ensure that homozygotic mutations are not missed. These have been rare in the proximal promoter region, but they do exist and this mixing protocol is used successfully for their detection. As an example of the pattern complexity, representative DFPLC patterns are shown for the proximal promoter mutations in Figure 17.

## EXAMPLE 14

### Analysis Of Double Variants

Some subjects were found to have more than one polymorphism in the *CHRNA7* core promoter. To determine whether these were on the same chromosome, the two alleles were examined individually by cloning. The PCR products were generated with the GC-RICH PCR system (Roche), with final concentrations of 1X buffer, 2.0mM magnesium chloride, 0.25mM deoxynucleotide triphosphates, and 0.5  $\mu$ l of enzyme mix in a 25  $\mu$ l volume. Three primer sets were used (*See*, Table 6): primer set 1, core promoter to intron 1; primer set 2, core promoter only; and primer set 6, core promoter to intron 2. The reaction for the smaller product, generated with primer set 2 (0.8 $\mu$ M concentration of each primer), also included 1.0M GC-RICH resolution solution, while for the larger products generated with primer set 1 (0.4 $\mu$ M concentration of each primer) or primer set 6 (0.4 $\mu$ M concentration of each primer), 0.8M GC-RICH resolution solution was included. All PCR products were amplified in a Perkin-Elmer 480 PCR thermocycler by means of the following program: 96°C, 3 min; 33 cycles at 96°C, 30 sec; 56°C, 30 sec; 72°C, 7 minutes. The appropriate PCR bands were gel-purified with the cONCERT Rapid Gel Extraction System (Life Technologies), and cloned into the PCR 4-TOPO vector with the TOPO TA Cloning Kit (Invitrogen). Plasmid DNA was isolated with the S.N.A.P. Miniprep kit (Invitrogen) and analyzed by DNA sequencing. Approximately 20 clones were sequenced for each double variant cloned.

In the individuals included in this study, 8 doubly polymorphic patterns were found (subjects had more than 1 polymorphism in the core  $\alpha 7$  promoter. Five of these double variants were found only in schizophrenic patients (marked with asterisks in Table 8). DNA fragments were cloned and sequenced from individuals with most of the double variant patterns isolated thus far. Three primer sets of Table 6 were used: 1 that amplified the core promoter of 271 bp (primer set 2); another set that amplified the core promoter, exon 1, and part of intron 1 (primer set 1); and a primer set that amplified the core promoter, exon 1, intron 1, exon 2, and part of intron 2 (primer set 6). Two variants were never found on the same chromosome, and only 2 alleles were present in all cases examined, indicating that the core promoter region is not duplicated in these individuals and, further, that each variant is a separate allele. Thus, polymorphisms in the core promoter of the full-length  $\alpha 7$  nicotinic receptor gene are found more

frequently in schizophrenic individuals than in subjects with no family history of schizophrenia, and double variants are likely to result from inheritance of one mutant allele from each parent.

## EXAMPLE 15

### Analysis Of Promoter Function

Promoter function was determined by means of a luciferase reporter gene assay. To identify a core promoter sequence in the 5' sequence upstream of the ATG translation start site in the  $\alpha 7$  nicotinic receptor gene, fragments of this region were subcloned into the pGL3-Basic Vector (Promega), using PCR and the pGEM-T Easy Vector System II kit (Promega). Initially, a 2602 bp fragment was inserted into the pGL3 vector (See, Figure 12 panel A, -2600 to +2). A 1064 bp clone was generated by partial *Pst*I digestion of the original fragment and cloned into the pGL3-Basic Vector. *Pst*I was then used to subclone a fragment of 231 bp, containing the proximal promoter region, which is conserved in the bovine  $\alpha 7$  gene (Carrasco-Serrano *et al.*, J Biol Chem, 273:20021-20028, 1998). Transcription factor consensus sequences in the 5' upstream region were identified with the TRANSFAC program available on the internet, courtesy of the Research Group Bioinformatics/AG Bioinformatik. Variants discovered in the mutation screen were introduced into the normal *Pst*I core promoter clone by using the QuikChange Site-Directed Mutagenesis Kit (Stratagene). Transient transfections were done with ProFection Mammalian Transfection Calcium Phosphate System (Promega), with the human neuroblastoma cell line SHSY-5Y (Flora *et al.*, Eur J Pharmacol, 393:85-95, 2000). The SHSY-FY cell line was grown in 1:1 Ham F12:DMEM, and 10% fetal calf serum, plated at  $2 \times 10^5$  cells/35mm plate. Five  $\mu$ g plasmid DNA prepared with EndoFree Plasmid Kits (Qiagen) was cotransfected with 1  $\mu$ g of pRL-TK Vector (Promega). Cells were harvested after 48 hours and luciferase activity was measured with the Dual-Luciferase Reporter Assay System (Promega) and a Turner Designs Luminometer Model TD 20/20.

*In vitro* functional analysis was performed for several of the polymorphisms found more frequently in schizophrenic subjects. A luciferase reporter gene assay was used to compare the normal core promoter sequence with a fragment containing one of these variants. As shown in Figure 13, variants at -86 bp, -92 bp, -143 bp, -178 bp, -194 bp, and -241 bp decreased transcription of the luciferase reporter gene in this *in vitro* assay, indicating that presence of one of these polymorphisms in the core promoter region decreases transcription from the gene. The

–86 bp C/T variant resulted in a decrease in luciferase transcription of 20% ( $P < 0.0001$ ). The functional promoter mutations examined thus far were statistically more prevalent ( $\chi^2_{df=1} = 7.302$ ,  $P = 0.007$ ) in schizophrenic patients than in the control subjects.

## EXAMPLE 16

### Statistical Analysis

For the statistical analysis, total counts from schizophrenic individuals included polymorphisms detected in only one schizophrenic individual per family, unless a second mutation was also present in another affected individual. In this case, the second variant was also counted. Subjects homozygous for the common allele were also counted. This strategy was chosen to report the full range of polymorphisms in schizophrenic patients without biasing the results by including multiple individuals who have the same polymorphism based on common ancestry.  $t$  tests were used to compare means. A Satterthwaite  $t$  test was used for comparison of means with difference variances; chi-squared tests and logistic regression were used to compare prevalence rates. For the double variants in the promoter region, cloning experiments indicated that each polymorphism is a separate allele.

Although promoter variants were found in control subjects, they were fewer in number than in schizophrenic patients. In complex disorders where multiple gene variants may be interacting with environmental factors to produce the disease, it has been suggested that functional polymorphisms are likely to be common in the general population, where each may have a more elementary phenotype, such as a biochemical or electrophysiologic abnormality that is part of the pathophysiology of the illness (Lander and Schork, *Science*, 265:2037-2048, 2000; and Gershon, *Biol Psychiatry*, 45:551-558, 1999). The association of CHRNA7 promoter polymorphisms was examined, in the living control subjects, with a functional electrophysiologic assay (*e.g.*, inhibition of the P50 response to paired auditory stimuli). The P50 auditory sensory gating was measured in 151 of the 152 live control subjects examined in this study. The range of P50 ratios (T/C) for controls was 0.00 to 1.91. Overall mean P50 ratio was  $0.22 \pm 0.27$ . There were 38 adult schizophrenic subjects examined locally where P50 recording was done. The mean P50 ratio for these patients was  $0.92 \pm 1.02$ , with a range of 0.00 to 4.96. Eighteen patients with childhood-onset schizophrenia, included in the mutation screen, also had their P50 recorded. Their mean was  $1.05 \pm 0.91$ , similar to that of the adult patients with schizophrenia.

The mean ratio for the schizophrenic patients was significantly greater than that of the control subjects ( $t_{205}=8.49$ ,  $P<0.0001$ ).

Tracings for subjects with and without the -86 bp C/T polymorphism are shown in Figure 14. A control subject with the normal C/C genotype had a T/C ratio (P50 ratio) of 0.14, indicating that the test response to the second auditory stimulus was being inhibited. However, a control subject carrying a -86 bp C/T heterozygotic genotype had a T/C ratio of 0.60, demonstrating a much lower level of inhibition. A schizophrenic patient with the -86 bp C/T genotype also had a higher T/C ratio of 0.54. These results indicate that the presence of a promoter variant is associated with decreased inhibition in the sensory gating paradigm and, hence, a higher T/C (P50) ratio.

The relationship between the means for the P50 T/C ratios and the presence of *CHRNA7* promoter variants was then examined in the 151 control subjects. As shown in Figure 15 panel A, the mean P50 ratio for controls with no *CHRNA7* promoter variant was  $0.179 \pm 0.014$ . However, the mean for control subjects with one of the single or double variants was  $0.458 \pm 0.055$ . The results were analyzed, using a Satterthwaite  $t$  test for samples with different variances. The control subjects with no polymorphisms had a significantly lower mean P50 ratio than control subjects in whom a promoter variant was found ( $P<0.0001$ ), demonstrating a strong relationship between the presence of a promoter variant and decreased sensory processing.

In the patients with adult-onset disease, where P50 had been recorded, 7 of 8 polymorphisms in the core promoter were found in schizophrenic patients with P50 ratios greater than 0.50. In the 18 patients with childhood-onset disease, wherein P50 had been recorded there were 7 polymorphisms, 5 of which were found in subjects with P50 ratios greater than 0.50. These results indicate that a similar relationship between the *CHRNA7* promoter polymorphisms and the recorded P50 ratio exist in both adult-onset and childhood-onset schizophrenia.

Logistic regression analysis of the control data indicated, that the presence of promoter variants is better described by three groups, than by a regression line on the P50 range (See, Figure 15, panel B). One group with individual average P50 ratios less than 0.20 was found to have stable auditory gating. A second group with P50 ratios between 0.20 and 0.50 was found to have a less stable filtering mechanism. A third group with P50 ratios consistently greater than 0.50 was found to exhibit very little auditory gating, similar to what has been described herein in the schizophrenic population. Control subjects with no polymorphism in the core  $\alpha 7$  promoter were

found to have P50 ratios in the first 2 groups, with most in the less than 0.20 group. In contrast, controls with polymorphisms were more evenly distributed among the three P50 groups, while only subjects with a promoter variant were found to have P50 ratios greater than 0.50.

5

## EXAMPLE 17

### Subject Selection and Sample Collection for *CHRNA7* and *dupCHRNA7* Analysis

Samples from 171 families with schizophrenic members and 185 samples from controls were available for screening. The sample population included 86 families from the NIMH Schizophrenia Genetics Initiative. Sixteen of these families had been used in a sib pair analysis showing greater than 50% inheritance-by-descent to a dinucleotide marker D15S1360 in the *CHRNA7* gene (0.58;  $P < 0.0024$ ) as described (Leonard *et al.*, Am J Med Genet, 81:308-312, 1998). Nine probands from the P50 linkage analysis (Freedman *et al.*, Proc Natl Acad Sci USA, 94:587-592, 1997) were also included and the remaining samples were collected in the Denver Schizophrenia Center. When postmortem brain samples were used, diagnosis was based upon review of medical records and family and physician interviews. Of the controls, 166 were interviewed and found to have no evidence for current or past psychosis, using two different interviews (See, First *et al.*, Structured Clinical Interview for DSM-IV Axis I Disorders-Non-Patient Edition, SCID-I/NP, version 2.0, NY:Biometrics Research Department, New York State Psychiatric Institute, 1996; and Endicott *et al.*, Family History - Research Diagnostic Criteria interview, FH-RDC, 3rd edition, NY:Research Assessment and Training Unit, New York State Psychiatric Institute, 1978). In addition, auditory evoked potentials were recorded on controls, using published methods (Freedman *et al.*, Schiz Res, 4:233-243, 1991).

The 84 samples used for cDNA mapping of the eight common variants were collected in the Denver Schizophrenic Center, and were primarily obtained from Caucasian subjects. As shown in Table 15, 28 samples from Caucasian individuals with schizophrenia (18 male and 10 female) and 49 samples from Caucasian controls (35 male and 14 female) were analyzed.

25

**Table 15. Subjects Used For Screening the *CHRNA7* and *dupCHRNA7* Genes**

<b>Subjects</b>	<b>Schizophrenics</b>			<b>Controls</b>		
<b>Ethnicity</b>	<b>Total</b>	<b>white</b>	<b>Black</b>	<b>Total</b>	<b>white</b>	<b>Black</b>
tissue						
brain	28	25	3	20	19	1
lymphocytes	4	3	1	32	30	2
totals	32	28	4	52	49	3

**Table 16. PCR Primers for Amplification of the *CHRNA7* Gene**

Product	Variants	SEQ ID NO	Strand	Primers	T <sub>A</sub>	SSCP
Exon 1	45, +82	137	S	GCGGCGAGGTGCCTCTGT	60°C	25°C
		138	AS	GGATCCCACGGAGGAGTGGAG		
Exon 2		139	S	CCTGCCCCGGGTCTTCTCTCCT	58°C	25°C
		140	AS	AACTAGAGTGCCCCAGCCGAGCT		
Exon 3		141	S	AACAACGCTCTCGACAGTCAGATC	58°C	25°C
		142	AS	AAGATCTTGCAGCCCATGGGAG		
Exon 4	334	143	S	GGAATTCTCTTTGGTTTGCAC	58°C	6°C
		144	AS	ACATATCCAGCATCTCTGTGA		
Exon 5	370	145	S	TCATGCAGTCCTTTTCCTGTTTC	60°C	6°C
		146	AS	CTCGCTTCAGTTTTCTAACATGG		
Exon 6		147	S	GGAAGTGTGTGTATTTTCAGC	58°C	both
		148	AS	TTAAAGCTTGCCAGGAATAGG		
Exon 7		149	S	GCTTGTGTGTGGTATACACATTG	58°C	both
		150	AS	TCCAGAGCTGATCTCAGCAGAAG		
Exon 8	861	151	S	GAGGAACCGCTGTGTGTTTAT	58°C	25°C
		152	AS	CTGGGCACACTCTAACCCTAACC		
Exon 9		153	S	TGTGACGTGCAGTGCCACAGGA	60°C	25°C
		154	AS	AAACCCTAGGAGGAGCCTCCTT		
Exon 10		155	S	GATCAGCCCGTTTCCGCCTCAG	58°C	both
		156	AS	CCGATGTACAGCAGGTTCCCGTTGC		
Exon 6*	497-8	157	S	CAGTACCTGCCTCCAGG	58°C	25°C
		158	AS	TCCAAGGACCAGCCTCCGTAAGA		
Exon 7*	654/690	159	S	CTATGAGTGCTGCAAAGA	58°C	25°C
		160	AS	CAGGGGATCAGCAGGTT		
Exon 7*	698/+21	161	S	GCCGCAGGACACTCTAC	58°C	25°C
		162	AS	TCCAGAGCTGATCTCAGCAGAAG		
Intron 7*	-11, -20, -29	163	S	GCCCCCTCGTTAGACAGAATTGAG	58°C	25°C
		164	AS	CTGGGCACACTCTAACCCTAACC		
Exon 10*	1044, 1116	165	S	GATCAGCCCGTTTCCGCCTCAG	58°C	25°C
		166	AS	CCGATGTACAGCAGGTTCCCGTTGC		
Exon 10*	1335	167	S	TCCCGACCCCCGACTCT	58°C	6°C
		168	AS	TGATGGTGAAGACCGAGAAGG		
Exon 10*	1269, 1354, 1456	169	S	TCCCGACCCCCGACTCT	58°C	25°C
		170	AS	TGATGGTGAAGACCGAGAAGG		
Exon 10*	1466	171	S	CCTTCTCGGTCTTCACCATC	58°C	25°C
		172	AS	GCCTCCACGAAGTTGGGAGC		
Exon 10*	1487	173	S	GGTCCGCTACATTGCCAA	58°C	25°C
		174	AS	CCTTGCCCATCTGTGAGTT		
3'UT*	1737, 1837	175	S	GTGTTGCTTACGGTTTCTT	58°C	25°C
		176	AS	TTTCAGGTAGACCTTCATGCAGACA		
cDNA#		177	S	TGCCCATCTGTGAGTTTCCACATG	72-68°C	
1-10		178	AS	CGCTGCAGCTCCGGGACTCAACATG		
cDNA#		179	S	CTCGGTGCCCTTGCCATTT	72-68°C	
D-10		180	AS	CCTTGCCCATCTGTGAGTTTCCAC		

## EXAMPLE 18

### Mutation Screening of the *CHRNA7* and *dupCHRNA7* Genes

5 All schizophrenic subjects in each family were screened for polymorphisms to detect the presence of different variants in related individuals. Initially, a strategy was used to screen genomic DNA from 96 samples from individuals where postmortem brain tissue or lymphoblasts were available. This was done because mRNA would be needed for the mapping of variants to either the full-length *CHRNA7* or its duplication (*dupCHRNA7*). In the initial gene mutation  
10 screen, all the exons, intron/exon boundaries, and the 3' untranslated region (UT) were examined by means of single-strand conformation polymorphism (SSCP) analysis using the primers shown for exons 1-10 in Table 16. Exon 10 and the 3' UTR were divided into an additional eight overlapping PCR fragments of approximately 200 bp, designed from the *CHRNA7* sequence (GenBank Accession No. U40583). For SSCP analysis the primer sets were kinased using [ $\gamma$ -  
15  $^{33}\text{P}$ ] ATP with Promega T4 kinase, then used to amplify regions of the *CHRNA7* gene by PCR. Briefly, PCR was done using *Taq* Gold™ and GeneAmp® PCR System 9600 (Perkin-Elmer, Foster City, CA) with the following program: 95°C for 3 min; then 35 cycles of 95°C for 30 sec, 58°C for 30 sec, and 72°C for 30 sec; followed by 72°C for 3 min. Specific annealing temperatures ( $T_A$ ) are provided in Table 16). The products, amplified and analyzed separately,  
20 were denatured with loading dye (7.26 M urea, 60% formamide, 22 mM EDTA, 32 mM NaOH, 0.25% bromophenol blue, 0.25% xylene cyanol), and separated on GeneAmp detection gels (Perkin-Elmer) run at 25°C and 6°C using a BioRad Power Pac 3000 with a temperature probe. Samples with unique SSCP patterns were sequenced and polymorphisms were correlated with the SSCP patterns. Identified variants were subsequently screened in additional genomic  
25 samples from controls, individuals with schizophrenia, and family members, using the appropriate primers and gel conditions. In Table 16, additional primer sets used to detect specific variants are indicated with an asterisk, while primers used in primary RT-PCR for mapping are indicated with a pound sign.

Mutation analysis of the  $\alpha 7$  nicotinic receptor gene *CHRNA7* and its partial duplication  
30 *dupCHRNA7* was carried out using SSCP, and sequence analyses. Figure 16 panel A depicts the 15q13-q14 region containing *CHRNA7* and *dupCHRNA7*. The unique dinucleotide marker

D15S1360, used in several linkage studies (Freedman *et al.*, Am J Med Gen, 105:794-800, 2001; Freedman *et al.*, Proc Natl Acad Sci USA, 94:587-592, 1997; Leonard *et al.*, Am J Med Genet, 81:308-312, 1998; and Freedman *et al.*, Am J Med Gen, 205:20-22, 2001), lies in intron two of *CHRNA7* (Leonard *et al.*, Arch Gen Psychiatry, 59:1085-1096, 2002). D15S1031 and D15S144, (also single copy) flank the full-length *CHRNA7* gene and duplicated cassette (duplicon). Unique loci D15S1043 and D15S165 flank the proximal duplicon. The duplicon contains exons 5-10 of the *CHRNA7* gene, the dinucleotide repeat L76630, exons D'-D-C-B-A, and the Expressed Sequence Tag (EST) WI13983. The transcripts from both  $\alpha 7$  containing genes are shown in Figure 16 panel B, with their unique 5' ends and the number of variants mapped to each exon. The orientation of the duplicon is shown as head to tail, determined from yeast artificial chromosome (YAC) mapping from two separate YAC libraries (Gault *et al.*, Genomics, 79:197-209, 2002). A head to head orientation has been reported based on BAC clone mapping from a single library (Riley *et al.*, Genomics, 79:197-209, 2002), suggesting that the orientation of this duplicon may be polymorphic.

Thirty-three variants in the *CHRNA7* gene cluster were identified in genomic DNA from individuals with schizophrenia and controls of Caucasian, African American and Hispanic descent (Tables 17, 18 and 19). Twenty-one different variants were found in the coding region of the  $\alpha 7$  genes, including 10 non-synonymous variants. Base pair numbering is from the first base pair in exon 1. Allele frequencies for 14 of the rare variants were calculated and are shown in Table 20. Allele frequencies for the more common variants were not determined because they could be homozygous in either *dupCHRNA7* or full-length *CHRNA7* genes. Six variants were found more frequently in the African Americans than the Caucasians (Table 21). Three variants at: 497-8 bp (2bp deletion), 654 bp, and 1466 bp, were found more frequently in Caucasians than in African Americans. Two rare, but non-synonymous variants in Exon 5 at 370 bp, and in Exon 7 at 698 bp, were found only in Hispanics (See, Table 17).

In Tables 17-21, the following nomenclature and abbreviations apply: E, exon; I, intron; V, number of individuals with the variant; T, total number of individuals;  $\alpha 7$ , full length gene; and  $d\alpha 7$ , duplicate gene. Numbering for exons and 3'UT is from the ATG start, while numbering for introns is from the 5' donor splice site (+) or 3' acceptor splice site (-). Variants: <sup>a</sup>exon 4, I112V; <sup>b</sup>exon 5, A124T; <sup>c</sup>exon 6, 2 bp deletion at L166 > in Caucasian subjects  $X^2 = 48.66, 1, P < 0.0001$ ; <sup>d</sup>exon 7, Y233C; <sup>e</sup>exon 9, G324R; <sup>f</sup>exon 10, S372R; <sup>g</sup>exon 10, E452K; <sup>h</sup>exon

10, I486V; <sup>i</sup>exon 10, S489L; <sup>j</sup>exon 10, A496D; <sup>k</sup>intron 9,  $X^2 = 9.986, 1, P=0.0016$ ; and <sup>l</sup>provisional mapping.

**Table 17. Non-synonomous Variants Identified in *CHRNA7* and *dupCHRNA7***

Site	bp	change	Schizophrenics						Controls						Map	
			Cauc.		Afr. Am.		Hisp.		Cauc.		Afr. Am.		Hisp.			
			V	T	V	T	V	T	V	T	V	T	V	T	V	T
E4 <sup>a</sup>	334	A→G	0	113	1	43	0	6	0	103	0	55	1	8	X	
E5 <sup>b</sup>	370	G→A	0	112	0	42	1	7	0	100	0	53	2	8		X <sup>l</sup>
E6 <sup>c</sup>	497/8	–TG	68	86 <sup>c</sup>	15	50	5	7	48	71 <sup>c</sup>	12	54	4	4		X
E7 <sup>d</sup>	698	A→G	0	85	0	38	1	7	0	58	0	4	0	4		X <sup>l</sup>
E9 <sup>e</sup>	970	G→A	0	110	10	52	0	6	0	79	4	52	0	7		X
E10 <sup>f</sup>	1116	C→G	0	106	0	36	0	6	0	71	1	49	0	4		X
E10 <sup>g</sup>	1354	G→A	1	102	0	41	0	6	1	63	0	3	0	3	X	
E10 <sup>h</sup>	1456	A→G	0	91	0	40	0	6	1	58	0	4	0	3		X
E10 <sup>i</sup>	1466	C→T	23	110	7	49	1	7	27	82	3	52	3	7		X
E10 <sup>j</sup>	1487	C→A	0	62	1	10	0	6	0	12	0	50	0	3	X	

5

**Table 18. Synonomous Variants Identified in *CHRNA7* and *dupCHRNA7***

Site	bp	change	Schizophrenics						Controls						Map	
			Cauc.		Afr. Am.		Hisp.		Cauc.		Afr. Am.		Hisp.			
			V	T	V	T	V	T	V	T	V	T	V	T	V	T
E1	45	G→A	1	99	0	41	0	6	0	64	0	3	0	3	X	
E7	654	C→T	77	90	32	47	5	6	57	70	3	4	3	3		X
E7	690	G→A	82	83	36	36	6	6	59	59	4	4	3	3	X	X
E8	861	C→T	4	98	1	40	1	7	1	59	0	4	0	3		X
E9	921	G→A	2	112	1	45	0	6	4	77	0	50	0	7		X
E9	933	G→A	56	127	28	53	6	8	39	79	18	50	6	7	X	
		A only	2	127												
E9	966	C→T	1	110	6	46	0	6	0	79	2	52	0	7		X <sup>l</sup>
E10	1044	C→T	12	123	3	43	0	6	9	72	1	55	1	5		X
E10	1116	C→T	2	107	8	44	0	6	0	71	6	54	1	5	-	-
E10	1269	C→T	75	95	29	40	5	6	47	57	2	3	3	3	X	X
		T only	2	95	1	40			1	57						
E10	1335	C→T	32	74	3	11	2	7	30	65	2	4	1	3	X	X

**Table 19. Non-coding Variants Identified in *CHRNA7* and *dupCHRNA7***

Site	bp	change	Schizophrenics						Controls						Map	
			Cauc.		Afr. Am.		Hisp.		Cauc.		Afr. Am.		Hisp.			
			V	T	V	T	V	T	V	T	V	T	V	T	V	T
I2	+75	G→A	0	87	1	38	0	6	1	50	0	3	0	1	X	
I2	+82	A→C	0	87	2	38	0	6	0	50	0	3	0	1	X	
I3	-9	A→G	0	113	3	45	0	6	0	103	1	55	0	8	X	
I7	+21	C→T	21	31	1	6	1	3	3	3	0	0	0	0	-	-
I7	-11	+GTT	10	38	2	4	0	4	5	10	0	1	0	1	-	-
I7	-20	G→A	15	37	1	4	2	4	5	10	1	1	1	1	-	-
I7	-29	T→G	1	37	0	4	0	4	0	10	0	1	0	1	-	-
I9	+19	C→T	0	43	0	5	0	7	1	78	4	54	0	7	-	-
I9	+27	-TCGGAG	0	110	1	44	0	6	0	78	2	54	0	7	-	-
I9	+37	G→C	56	126	36	58 <sup>k</sup>	6	8	38	79	17	53	6	7	-	-
3'UT	1737	C→A	1	34	0	5	0	2	0	33	0	1	0	1	-	-
3'UT	1837	T→G	0	34	1	6	0	2	0	33	0	1	0	1	-	-

**Table 20. Allele Frequencies of Rare Variants**

Site	bp	change	Amino Acid	Ethnicity	Frequency Schizo.	Frequency Controls	Map	
							$\alpha 7$	$d\alpha 7$
E1	45	G→A		Cauc.	0.005	0	X	
I2	+75	G→A		Cauc.	0	0.01	X	
I2	+82	A→C		Afr. Am.	0.026	0	X	
I3	-9	A→G		Afr. Am.	0.033	0.009	X	
E4	334	A→G	I112V	Afr. Am./Hisp.	0.010	0.008*	X	
E8	861	C→T		Cauc.	0.020	0.008		X
E9	921	G→A		Cauc.	0.009	0.026		X
E9	921	G→A		Afr. Am.	0.011	0		X
I9	+19	C→T		Cauc.	0	0.006	-	-
I9	+27	-TCGGAG		Afr. Am.	0.011	0.019	-	-
E10	1116	C→G	S372R	Afr. Am.	0	0.010	-	X
E10	1354	G→A	452K	Cauc.	0.005	0	X	-
E10	1456	A→G	I486V	Cauc.	0	0.009		X
3'UT	1737	C→A		Cauc.	0.015	0	-	-

\*Variant 334 was found in 1/43 African American and 0/6 Hispanic schizophrenics, and 0/55 African American and 1/8 Hispanic controls.

**Table 21. Variants with Significantly Different Frequencies by Ethnicity**

Site	bp	change	Caucasian		African American		P values	Map	
			V	T	V	T		$\alpha 7$	$d\alpha 7$
I3	-9	A→G	0	216	4	100	0.0104	X	
E6	497/8	-TG	116	167	27	104	<0.0001		X
E7	654	C→T	134	160	35	51	<0.0001		X
E9	966	C→T	1	189	8	98	0.0010		X
E9	970	G→A	0	189	14	104	<0.0001		X
I9	+19	C→T	0	121	4	59	0.0475	-	-
I9	+27	-TCGGAG	0	188	3	98	0.0394	-	-
E10	1116	C→T	2	178	14	98	<0.0001	-	-
E10	1466	C→T	50	192	10	101	0.0011		X

#### EXAMPLE 19

##### Mapping Variants to the *CHRNA7* and *dupCHRNA7* Genes

*CHRNA7* exons 5-10 are duplicated and nearly homologous (>99%), complicating the mutation screen. However, the duplicated exons are transcribed with different 5' sequence and thus were isolated as unique mRNA species. The cDNA primer sets, used to specifically amplify full-length cDNA from either *CHRNA7* or its duplication (*dupCHRNA7*), are listed as the last two entries in Table 16. These cDNA templates were then used to map the variants in exons 5-10, using RT-PCR and subsequent SSCP and sequence analysis of the RT-PCR products.

Eighty-four samples from the mutation screening study were used for cDNA mapping of the eight common variants. Immortalized cell lines were not available from the NIMH Schizophrenia Initiative samples and, thus, postmortem brain and immortalized lymphoblasts collected locally in the Denver Schizophrenia Center were utilized.

Immortalized lymphocytes were cultured 6 hours with 1 mg/ml cyclohexamide before RNA isolation. Total RNA was isolated from postmortem human hippocampus or cyclohexamide-treated immortalized lymphocytes, using TRIzol reagent (Life Technologies, Gibco-BRL). RNA was reverse transcribed (500 ng) using Superscript II reverse transcriptase components (Gibco-BRL) with 8  $\mu$ M random hexamers (Pharmacia & Upjohn Diagnostics, Kalamazoo, MI) and 0.5U placental RNase inhibitor (Boehringer-Mannheim, Indianapolis, IN). A primary PCR was performed using specific primers designed with Oligo software 4.1

(National Biosciences, Inc., Plymouth MN). Full-length *CHRNA7* transcripts were amplified using 1 M GC-melt and 10X cDNA buffer (Clontech, K1905-1) from the Advantage cDNA PCR kit (CLONTECH, Laboratories Inc., Palo Alto, CA) and a two-step program with annealing temperatures from 72°C to 68°C. Partially duplicated *dupCHRNA7* transcripts were amplified using 1 M GC melt and 5X cDNA buffer from the Advantage-GC cDNA PCR kit (K1907-1). These primary reactions were then analyzed using SSCP and sequence analysis.

As shown herein, exons 5-10 of the  $\alpha 7$  nicotinic receptor subunit gene are duplicated. Genomic variants in these exons, therefore, are contemplated to be present in either the full-length *CHRNA7* gene or in *dupCHRNA7*. Polymorphisms were mapped, when possible, to one of the two duplicons, utilizing mRNA isolated from either immortalized lymphoblasts or postmortem brain by using gene specific PCR. In some cases a given variant was present in both duplicons. In others, only tissue from a schizophrenic subject was available for mapping. In this case, the map site is indicated as provisional, since gene rearrangements or conversions could have occurred.

Eight of the more common variants were mapped in 32 samples from individuals with schizophrenia and 52 samples from control individuals (total of 84). Four common variants: the 497/8 2bp deletion, the neutral variant at 654 bp, the neutral variant at 1044 bp, and the amino acid changing variant at 1466 bp, all mapped only to *dupCHRNA7* (See, Table 22). The 2 bp deletion in exon 6 was found in 15 out of the 32 Caucasians with schizophrenia, 29 out of the 49 Caucasian control samples, 1 out of the 4 African Americans with schizophrenia, and 2 out of the 3 African American controls.

Three common neutral variants, at 690 bp, 1269 bp, and 1335 bp mapped to both duplicons. The very common variant at 690 bp mapped primarily to the duplicated gene (69 out of 72 individuals). The 1269 bp variant mapped to both *CHRNA7* genes in 14 out of 54 individuals, while the neutral variant at 1335 bp mapped primarily to the full-length *CHRNA7* gene, and the variant at 933 bp mapped only to the full-length gene. Variant 933 bp G→A is in linkage disequilibrium with an intronic variant, and is contemplated to involve splicing.

Table 22.

## Mapping of common variants identified in the CHRNA7 gene and its partial duplication

Exon/ Intron	bp	$\Delta$	Schizophrenics										Controls										Map	
			Caucasian					Af. Am.					Caucasian					Af. Am.					$\alpha 7$	$\alpha 7$
			genomic		cDNA		V	T	I	genomic		cDNA		V	T	I	genomic		cDNA					
			$\alpha 7$	$\Delta \alpha 7$	$\alpha 7$	$\Delta \alpha 7$				$\alpha 7$	$\Delta \alpha 7$	$\alpha 7$	$\Delta \alpha 7$											
																	$\alpha 7$	$\Delta \alpha 7$	$\alpha 7$	$\Delta \alpha 7$	$\alpha 7$	$\Delta \alpha 7$		
E-6 <sup>a</sup>	497-8	-TG	68	96	0	15	15	15	50	0	1	1	48	71	0	29	29	12	54	0	2	2		X
E-7	654	C→T	77	90	0	20	20	32	47	0	1	1	57	70	0	33	33	3	4	0	1	1		X
E-7	690	G→A	82	83	4	22	24	36	36	1	3	4	59	59	4	41	41	4	4	0	3	3	X	X
E-9	933	G→A	56	127	12	0	12	28	53	3	0	3	39	79	15	0	15	18	50	1	0	1	X	
		A only	2	127																				
E-10	1044	C→T	12	123	0	3	3	3	43	0	1	1	9	72	0	3	3	1	55	0	0	0		X
E-10	1269	C→T	75	95	9	13	18	29	40	2	3	4	47	57	18	22	31	2	3	0	1	1	X	X
		T only	2	95				1	40				1	57										
E-10	1335	C→T	32	74	6	2	7	3	11	0	0	0	30	65	17	3	19	2	4	2	0	2	X	X
E-10 <sup>b</sup>	1466	C→T	23	110	0	4	4	7	49	0	0	0	27	82	0	12	12	3	52	0	0	0		X

E, exon; I, intron. Numbering for exons and 3'UT is from the ATG start.  $\alpha 7$ , full-length gene.  $\alpha 7$ , duplicated gene. total, number of individuals with the variant that were mapped. <sup>a</sup> exon 6, 2bp deletion at leu a.a. 166; <sup>b</sup> exon 10, ser to leu a.a. 489.

Ten of the thirty-three variants in Tables 17-19 were not mapped. Seven of the unmapped variants lie in introns and could not be mapped using the cDNA specific RT-PCR methodology. One unmapped variant in exon 10 was discovered late in the screen and was found to be synonymous.

A large number of variants (12) were found in a short proximal promoter region 5' of the translation start as shown in Figure 16 panel A and as published (Leonard *et al.*, Arch Gen Psychiatry, 59:1085-1096, 2002). As described herein, many of the variants were found to functionally reduce transcription in a reporter gene assay and to be associated with both the P50 auditory gating deficit and with schizophrenia. The relationships of these promoter polymorphisms to some of the variants in the coding and non-coding sequence are discussed below.

### Non-synonymous Variants

The coding region of the full-length *CHRNA7* gene consists of 10 exons. Eleven variants mapping to the full-length gene are reported in Tables 17-19, three of which are non-synonymous. The A→G variant at 334 bp in exon 4, results in a conservative amino acid change of an isoleucine to a valine at amino acid 112. However, this residue lies in the putative agonist-binding site (Galzi *et al.*, Annu Rev Pharmacol, 31:37-72, 1991), where a conformational alteration is contemplated to result in a change in agonist affinity. The rare variant at 334bp was found in one African American schizophrenic but not in an affected sibling and in one Hispanic control subject. The control subject exhibited abnormal auditory evoked potential responses, having a P50 test to conditioning ratio of 1.91. Both subjects with this rare 334 bp variant also have a rare insertion in the  $\alpha 7$  core promoter (-190 +G), indicating that this represents a minor haplotype. The schizophrenic, however, also carries a core promoter mutation on the other chromosome (-178 -G).

The G→A variant at 1354 bp in exon 10 changes a glutamic acid to a lysine in the large intracellular loop of the protein. A glutamic acid at this position is conserved across species. In the rat, a large deletion of sequence including this codon resulted in a two-fold increase of both  $\alpha$ -bungarotoxin binding and current in transfected oocytes (Valor *et al.*, Biochem, 41:7931-7938, 2002). However, the single non-conservative change from an acidic to a basic residue described

herein is expected to effect a functional change in the receptor. The rare variant at 1354 bp was found in one Caucasian schizophrenic and in one Caucasian control subject. Both of these subjects have normal core promoter sequences. Although not having the 1354 bp variant, an affected brother of the schizophrenic has a mutation in the core  $\alpha 7$  promoter (-86 bp), indicative of two  $\alpha 7$  alleles for schizophrenia in this family.

The C→A variant at base pair 1487 in exon 10 changes an alanine to an aspartic acid in the extracellular carboxyl terminus. The 1487 bp variant was found in one African American schizophrenic but not in an affected child. A family member with an abnormal P50 test to conditioning ratio of 61.7 carried an  $\alpha 7$  core promoter mutation (-191 G→A), again indicative of two alleles for schizophrenia.

Sixteen variants found in  $\alpha 7$  exons 5-10 mapped to the duplicated gene *dupCHRNA7*, which is also in the region of chromosome 15q14 genetically linked to schizophrenia (Tables 17-19 and Figure 16 panel A). The mRNA for *dupCHRNA7* is expressed in multiple tissues, including brain (Drebing *et al.*, Soc Neurosci Abst, 24:832, 1998). *DupCHRNA7* is present in only one copy in approximately 30% of the general population, but is homozygotically deleted in 5% of schizophrenic subjects (Gault *et al.*, Genomics 52: 173-185, 1998; and Leonard *et al.*, Biol Psych 49:571, 2001). Recent evidence suggests that *dupCHRNA7* transcripts are translated, but the function of this protein is not yet known (Lee *et al.*, Soc Neurosci Abst, 27:144.10, 2001). Six single nucleotide polymorphisms (SNP) change amino acids in a putative open reading frame found in *dupCHRNA7* (370 bp in exon 5, 698 bp in exon 7, 970 bp in exon 9, and 1116 bp, 1456 bp, and 1466 bp in exon 10).

A 2 bp deletion at bases 497/8 in exon 6 was found in one copy of the duplicated gene in 57.5% of schizophrenic subjects and in 49.6% of controls (not a significant difference). It was, however, found more frequently in Caucasian control subjects than in African American controls ( $X^2 = 25.31$ ,  $p < 0.0001$ ). This deletion, found only in *dupCHRNA7*, shifts the reading frame, resulting in three stop codons within the next 53 codons. These stop codons, however, are the most frequently skipped during translation (MacBeath and Kast, BioTechniques, 24:789-794, 1998). Further, the site surrounding the deletion in exon 6 is a consensus exon splice enhancer site (ESE) for enhancer factor SC35 (Cartegni *et al.*, Nat Rev Gen, 3:285-298, 2002). Deletion of the two base pairs (TG) is contemplated to disrupt this site, indicating that exon 6 is spliced out in these subjects, leaving an exon 5/exon 7 junction. This splice variant would leave the

coding sequence in frame. Deletion of exon 6 removes the cysteine bridge and part of a putative ligand binding site, leaving the remainder of the  $\alpha 7$  coding sequence intact. In the analysis of the *CHRNA7* proximal promoter described herein, subjects with a promoter variant were much less likely to have a 2bp deletion in exon 6 of the *dupCHRNA7* gene ( $X^2 = 16.46$ , 1;  $p < 0.0001$ ).

5 There was also a striking relationship with a three base pair insertion in intron 7. Every subject (50 out of 50) with a 2 bp deletion in exon 6 of the *dupCHRNA7* gene, also had this insertion (+GTT) at the -11 bp position in intron 7. This intronic variant is contemplated to reside in the gene duplication rather than in the full-length gene.

## 10 **Synonymous Variants**

Eleven SNPs in the coding regions that do not result in an amino acid change were found. Four conservative exon variants at bp 690, 1269, and 1335 map to both the duplicated gene and the full-length *CHRNA7* gene. The variant at 690 bp in exon 7 is the most common variant found in the  $\alpha 7$  nicotinic receptor genes, and it is heterozygous in genomic DNA from 190 of 15 191 samples examined. The G primarily maps to *CHRNA7* and the A primarily maps to *dupCHRNA7*. The 1269 bp and 1335 bp variants were found in 80% and 43% of all subjects, respectively.

Another common synonymous variant in exon 9, at bp 933, is of interest. It was found only in the full-length gene and is also inversely associated with the presence of a polymorphism 20 in the proximal promoter in all subjects examined,  $X^2 = 6.916$ , 1;  $p = 0.0085$ . The association was significant in the controls ( $X^2 = 5.183$ , 1;  $p = 0.0228$ ), but only suggestive in the schizophrenic subjects. The 933 G→C variant is found within the loop of a putative stem and loop structure formed by a tri-nucleotide repeat of (GGT)<sub>3</sub> and its complement repeat (ACC)<sub>3</sub> in exon 9 ( $\Delta G = -16.2$  kcal/mol). The 933 bp variant is also in linkage disequilibrium with a 25 common intronic variant in intron 9 as discussed below.

## **Intronic Variants**

Ten intron changes were identified, none of which change the consensus sequences at RNA splice junctions. However, a number of these variants may affect splicing by introducing a 30 favorable splice site or affecting the binding sites of splice enhancer proteins. In intron 3, a variant at -9 (G→A) changes the sequence near the 3' acceptor site to a sequence identical to

nine bp in exon 4, thereby forming a cryptic splice site. Although found in only 3 of 45 African American schizophrenic families (3/90 alleles), this polymorphism was found in only 1 of 55 African American controls (1/110 alleles). The single control subject with this variant had a P50 (test to conditioning ratio) of 0.32, in the unstable range, and had been diagnosed with major depression.

The intron 7 variant at -11 (+GTT) was mentioned above in relation to the 2 bp deletion in exon 6. Insertion of these three base pairs introduces additional pyrimidines into the splice acceptor site for exon 7, possibly increasing site use. Another intron 7 variant at -20 (G→A), is inversely associated with the presence of proximal promoter variants. Only 1 of 29 subjects with the polymorphism had a promoter mutation, while 20 of 58 subjects with the wild-type sequence had a promoter polymorphism ( $X^2 = 10.17$ , 1;  $p = 0.0014$ ).

A variant in intron 9 (+37, counted from the splice donor site) was found more frequently in the African American schizophrenic sample than in the control sample ( $X^2 = 9.986$ , 1;  $P=0.0016$ ). This same variant was not found at significantly different frequencies in the Caucasian schizophrenia sample. One unaffected family member was identified with a homozygous C at base pair +37. Interestingly, the exon 9 variant at 933 bp (G→A) is in linkage disequilibrium with the intron 9 variant at +37 (G→C). Since the exon 9 variant at 933 appears to be in the full-length gene, it is contemplated that the intron 9 variant is also located in the full-length gene. If the intron 9 variant at +37 is not associated with the exon 9 variant at 933, then there is contemplated to be another polymorphism present nearby (e.g., exon 9 variant at 966, C→T), which is also present in the full-length gene. However, the exon 9 variant at 966 is rare and was only mapped in one individual who was a schizophrenic, and thus its map location is provisional.

## EXAMPLE 20

### Statistical Analysis of *CHRNA7* and *dupCHRNA7* Variants

Chi square statistics or Fisher's exact tests were used to determine whether a variant was found more frequently in the schizophrenic sample than the control sample. Allele frequencies were calculated for variants in exons 1-4, but could not be determined for polymorphisms in the duplicated exons. A case-control study was done. All schizophrenic subjects in each family were screened for polymorphisms to determine if variants cosegregate with affected family

members and to ensure that no mutations were missed. Total counts from schizophrenic individuals include one schizophrenic individual from each family, unless other schizophrenic family members differed from the proband at that nucleotide position. When this occurred, the other family member was also counted. The sample size provided sufficient power to detect a  
5 0.11 difference in allele frequency between the schizophrenic and control groups at a  $p < 0.05$  for an allele with a population frequency of 0.050.

Two population-specific loci, FY-null and RB2300, were used to estimate the degree of admixture in African American samples of schizophrenic individuals and controls (Parra *et al.*, Am J Hum Gen, 63:1839-1851, 1998). The FY-NULL\*1 allele is the normal allele with a C at -  
10 46 in the promoter of the DARC gene (Duffy antigen receptor of chemokines). The FY NULL\*1 allele has an allele frequency of 1.0 in European populations, 0 in African populations and 0.06-0.2 in African American populations (Parra *et al.*, *supra*, 1998). FY-NULL\*1 allele frequencies did not vary significantly between the African American controls (0.2) and the schizophrenic individuals (0.18) studied herein. The RB2300\*1 allele has an allele frequency of  
15 0.900 to 0.944 in African populations, 0.776 to 0.888 in African American populations, and 0.287 to 0.588 in European populations (Parra *et al.*, *supra*, 1998). The RB2300\*1 allele does not have a *Bam*HI polymorphism in intron 1 of the human retinoblastoma gene. The RB2300\*1 allele was found at a frequency of 0.82 in our African American controls and 0.86 in the African American subjects with schizophrenia (not significantly different). These data suggest that there  
20 is a similar degree of admixture in our African American control and schizophrenic samples and that differences in variant frequencies between these samples should not reflect ethnic bias.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the  
25 invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention, which are obvious to those skilled in molecular biology or  
30 related fields, are intended to be within the scope of the following claims.